

**THE ASSESSMENT OF ASSOCIATION BETWEEN INSOMNIA AND
RISK FACTORS IN THE PROVINCE OF NOVA SCOTIA**

by

VIKTORIA BASSARGUINA

B.Sc. (Hons.), Khabarovsk Technical University, 1995

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
MATHEMATICAL, COMPUTER, AND PHYSICAL SCIENCES
(MATHEMATICS)

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

March 2011

© Viktoria Bassarguina, 2011



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-75147-3
Our file *Notre référence*
ISBN: 978-0-494-75147-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

Scientific research on sleep problems reveals that insomnia has considerable impact on the daily functioning of the affected individual. This thesis investigates risk factors associated with insomnia in a random sample of individuals aged 12 years and older ($n = 5,018$) living in the province of Nova Scotia, Canada. A binomial logistic regression was used to assess the association between insomnia and health factors, demographic and socioeconomic characteristics, and lifestyle variables. Moreover, the bootstrap method was used to obtain estimates of the odds ratios, parameter estimates, and their standard errors in addition to the logistic regression.

The following findings are reported. Age less than 65 years, female gender, current level of smoking, arthritis, asthma, back problems, high blood pressure, bowel syndrome, diabetes, heart disease and migraine were associated with an increased risk of insomnia in the population of Nova Scotia. The bootstrap estimates for odds ratios were slightly higher than those obtained by the classical logistic regression model.

This is the first study conducted for insomnia in Nova Scotia for the years 2000-2010.

Table of Contents

Abstract.....	i
Table of Contents.....	ii
List of Tables	iv
List of Figures.....	v
List of Abbreviations	vi
Acknowledgements.....	viii
Dedication.....	ix
CHAPTER 1 – INTRODUCTION.....	1
1.1 Motivation and Contributions.....	3
1.2 Objectives	4
1.3 Thesis Overview	5
CHAPTER 2 – LITERATURE REVIEW	6
2.1 Defining Insomnia	6
2.2 Insomnia in the Context of Chronicity	8
2.3 Consequences of Insomnia	11
2.4 Factors Linked to Insomnia	13
2.5 Insomnia in Different Countries	17
2.6 Conclusions	21
CHAPTER 3 – METHODOLOGY	24
3.1 Source of the Data	24
3.2 Target Population.....	24
3.3 Design of CCHS 2007-2008.....	25
3.4 Statistical Analyses.....	26
3.5 Data Availability.....	27
3.6 Response Variable	27
3.7 Explanatory Variables	28
3.8 Logistic Regression Model.....	30
3.9 Bootstrap.....	32
3.10 Bootstrap Estimates	34

CHAPTER 4 – RESULTS.....	36
4.1 Data Screening and Cleaning.....	36
4.1.1 Missing Data.....	36
4.1.2 Analyzing Patterns of Missing Data	38
4.1.3 Univariate Outliers.....	42
4.1.4 Detecting Multivariate Outliers	43
4.1.5 Correlations.....	45
4.1.6 Multicollinearity	48
4.2 Logistic Regression Analysis	49
4.2.1 Prevalence of insomnia.....	49
4.2.2 Variable Selection.....	51
4.2.3 Final Multiple Logistic Regression Model.....	53
4.2.4 Comparing Models	59
4.3 Bootstrap.....	62
CHAPTER 5 – DISCUSSION AND CONCLUSIONS.....	71
5.1 Discussion.....	71
5.2 Conclusions	74
5.3 Limitations	74
5.4 Implications and Recommendations for Future Work.....	75
Bibliography	77
Appendix A – List of Variables in CCHS 2007-2008.....	83
Appendix B – Recoding Designations for the Variables in this Study.....	90
Appendix C – Histograms for the Lognormal and Gamma Distributions of the Study Variables	93

List of Tables

Table 1. Percentages of missing values.....	37
Table 2. The results of χ^2 test of independence	39
Table 3. Patterns of missing data	41
Table 4. Variable split in two different groups	43
Table 5. Spearman's rank correlation and p -values between INSOMNIA and explanatory variables	46
Table 6. Spearman's rank correlations between predictor variables.....	47
Table 7. Prevalence rates of insomnia in the sample	50
Table 8. Prevalence of medical problems	50
Table 9. Comparison results for the logistic regression models	52
Table 10. Logistic Regression Analysis of Insomnia.....	55
Table 11. Classification table	57
Table 12. Comparing models	60
Table 13. Odds Ratios for Univariate and Multivariate Logistic Regression Analyses	61
Table 14. Mean values of Odds Ratios, their confidence intervals and standard deviations for 10,000 bootstrap samples	63
Table 15. Standard errors for logistic regression parameter estimates and for bootstrapped parameter estimates	64
Table 16. Parameters of gamma distribution, its mean and variance.....	66
Table 17. Parameters of gamma distribution, its mean and variance.....	68
Table 18. Parameters of log-normal distribution, its mean and variance.....	68
Table 19. Recoding designations for the variables	90

List of Figures

Figure 1.	ROC curve for the final model	58
Figure 2.	Histogram for AGE 2 vs. 1	69
Figure 3.	Histogram for AGE 3 vs. 1	70
Figure 4.	Histogram for AGE 4 vs. 1	93
Figure 5.	Histogram for AGE 5 vs. 1	94
Figure 6.	Histogram for AGE 6 vs. 1	94
Figure 7.	Histogram for AGE 7 vs. 1	95
Figure 8.	Histogram for SMOKE 2 vs. 1	95
Figure 9.	Histogram for SMOKE 3 vs. 1	96
Figure 10.	Histogram for GENDER.....	96
Figure 11.	Histogram for ARTHRIT.....	97
Figure 12.	Histogram for ASTHMA	97
Figure 13.	Histogram for BACKPROB	98
Figure 14.	Histogram for BLOODPR	98
Figure 15.	Histogram for BOWEL.....	99
Figure 16.	Histogram for DIABETES	99
Figure 17.	Histogram for HEART.....	100
Figure 18.	Histogram for MIGRAINE.....	100

List of Abbreviations

AIC - Akaike Information Criterion

CAI - Computer Assisted Interviewing

CAPI - Computer Assisted Personal Interviewing

CATI - Computer Assisted Telephone Interviewing

CCHS - Canadian Community Health Survey

CI - Confidence Interval

DIS - Difficulty Initiating Sleep

DMS - Difficulty Maintaining Sleep

DSM-IV - Diagnostic and Statistical Manual of Mental Disorders

EDF - Empirical Distribution Function

ELSA - English Longitudinal Study of Ageing

GSS6 - Canadian General Social Survey Cycle 6

HALS - Health and Activity Lifestyle Survey

HR - Health Regions

HRQOL - Health-Related Quality of Life

LL - Log-likelihood

MAR - Missing at Random

MCAR - Missing Completely at Random

MNAR - Missing not at Random

MOS - Medical Outcomes Study

NLSAA - Nottingham Longitudinal Study of Activity and Ageing

NRS - Non-restorative Sleep

OR - Odds Ratio

RDD - Random Digit Dialing

SAS - Statistical Analysis Software

SC - Schwarz (Bayesian Information) Criterion

SES - Socioeconomic Status

SPC - Somatic and Psychological Complaint

URS - Unrestricted Random Sample

VIF - Variance Inflation Factor

Acknowledgements

I am thankful to my supervisor, Dr. Pranesh Kumar, for his encouragement, guidance, and support throughout this thesis. I have always been able to drop by at any time to get advice and help, which I greatly appreciate.

I would like to express my thanks to committee members: Dr. Peter MacMillan for his valuable guidance and advice on how to better organize my thesis and put it together, and Dr. Lee Keener for dedicating his time to provide beneficial feedback and suggestions. You both were very welcoming and available to help.

Additionally, I would like to thank Dr. Han Donker from the School of Business who agreed to serve as my external examiner despite the short notice.

Also, I would like to thank the University of Northern British Columbia and the Department of Mathematics in particular for the opportunity to carry out this research.

I would like to express my gratitude to my mother for her unconditional love, the ability to raise my spirit, and faith in me, although she is thousands miles away.

I am grateful to my friends, Alla and Egor, for their extended support and inspiration. With your help in endless ways you made this thesis possible.

Last but not least, I would like to thank Arber for using his editing skills and expertise to work through my thesis and helping me polish and tighten it. You have encouraged me by making me believe in myself and my abilities.

This thesis is dedicated to my adorable, extraordinary daughter, Dasha.

Darling, I love you deeply.

CHAPTER 1 – INTRODUCTION

According to scientific literature, an adult needs to have at least seven to eight hours of sleep every night in order to provide body and mind with the necessary rest. This implies that adults spend roughly one third of their lifetime sleeping, while the success and vitality of the remaining two-thirds of the lifetime get directly affected by that one-third fraction. The ability to solve problems, to perform tasks effectively and to concentrate on accomplishing them in time, and to make better life decisions increases when people get sufficient sleep. In addition, sleep influences mood, and constant lack of sleep can lead to depression. More importantly, people who do not get enough sleep or do not have good quality sleep are at risk of heart disease, high blood pressure, stroke, diabetes, and obesity, to name a few (Patlak, 2005).

In the literature, researchers define insomnia in different ways. In general, insomnia is defined as difficulty initiating and maintaining sleep or waking up too early (Patten, Choi, Gillin, & Pierce, 2000), as will be discussed in detail in Chapter 2. Insomnia can be categorized by duration: it is acute if it lasts for a fairly short period of time (less than 30 days), and chronic if it lasts more than 30 days (Turkosi, 2006). Experiencing life events, such as stressful or worrying situations (loss of a job, relationship conflict or simply anticipation of vacation or holiday), can initiate acute insomnia (National Sleep Foundation, 2010).

Insomnia affects daytime functioning and working abilities. Work performance, such as absenteeism, making errors at work, and difficulties accomplishing work assignments reflects the consequences of insomnia (Leger, Guilleminault, Bader, Levy, & Paillard, 2002).

Work productivity is diminished among people with insomnia due to work-related problems, i.e. higher rates of absenteeism, decreased concentration, and difficulty performing tasks (Okun et al., 2009; Roth, 2007). Impaired sleepers receive fewer job promotions, get paid less, and consequently perform less effectively. The loss of income due to insomnia affects those earning income. Also, “insomnia may influence a person’s decision to retire or enter the workplace” (Stoller, 2007).

Lack of sleep due to insomnia has a strong effect on driving performance. People deprived of sleep can easily get annoyed and angered, and, as a result, they have difficulty reacting, concentrating and paying attention to the road (Helpguide, 2010). Unsatisfactory sleep, sleep disorders, the length of restlessness, and extended work hours are all key causes of fatigue. Driver fatigue, which frequently causes drivers to fall asleep at the wheel, has been identified as the reason of many lethal road collisions (Akerstedt, 2000).

Approximately 400 Canadians die every year on the road due to driver fatigue (Highway Safety Roundtable, 2007). About 20% of Canadians, or an estimated 4.1 million people, confessed to have nodded off or fallen asleep on the road at least once in the past 12 months (Beirness, Simpson, & Desmond, 2005).

Shift work is also known to be related to an increase in sleep problems (Linton, 2004). Switching the patterns of daytime activity and nighttime rest frequently, leads an individual to experience symptoms of fatigue including, but not limited to, being forgetful, making poor decisions, having degraded alertness and slow reaction time, and as a result, reducing performance capability at work (Rosekind et al., 1996). Fatigue as a result of shift work has been related to many of the big accidents in modern times. For example, the Chernobyl incident was due to human error related to work scheduling (Mitler et al., 1988). Similarly, fatigue has been identified as a probable cause of the “Three Mile Island reactor

accident and the near miss incidents at the David [sic] Beese reactor in Ohio and at the Rancho Seco reactor in California” (Philip & Akerstedt, 2006), but also for the Disaster in Bhopal and the Space Shuttle explosion as well as the grounding of the oil tanker Exxon Valdez.

As these facts demonstrate, insomnia has a pervasive impact on the daily functioning of the affected individual. Thus, conducting research in order to investigate insomnia and related risk factors is a necessary task. It is a central objective of this thesis to carry out such research for the population of the province of Nova Scotia, Canada, using the Canadian Community Health Survey, 2007-2008. In what follows, the motivation, contributions and the objectives of this work are presented.

1.1 Motivation and Contributions

The Canadian Community Health Survey (CCHS) employed for this thesis is the survey covering several topics. The subject of sleep is an optional content that gets individually selected by health regions. This content in the CCHS was selected only by the province of Nova Scotia, suggesting that insomnia is a public health concern in this particular region. In comparison to other provinces, the province of Nova Scotia has the highest rates of deaths from cancer and respiratory disease, highest self-reported rates of arthritis and rheumatism, high rates of hospitalization for chronic diseases, second highest rates of deaths from breast cancer and diabetes, and a higher percentage of smokers than most other provinces (Capital Health Nova Scotia, 2008). These facts have brought about the need to conduct research for the province of Nova Scotia in order to investigate and determine possible links of health problems to insomnia.

The results presented in this study provide several contributions. First, risk factors for insomnia in the population of the province of Nova Scotia are determined. Second, health organizations in the province of Nova Scotia as well as other professional or academic organizations may find the results useful for taking measures to improve the health of Nova Scotians as well as for conducting research that is more specific to the complex issue of insomnia. Third, the results may help to raise public awareness of how crucial research on sleep and sleep disorders is. This may eventually raise funds for supporting sleep-related education and research. Also, at the government level, steps for reviewing policies and legislation may be undertaken to improve individual, family and community health. Growing recognition and understanding of the substantial consequences of insomnia, and sleep importance in general, support the rationale for doing research on sleep problems.

It is worth mentioning that the results presented in this thesis are the first reported on insomnia for the province of Nova Scotia for the years 2000-2010 using a population-based study.

1.2 Objectives

The primary objective of this research is to examine the association between insomnia and various health factors, demographic and socioeconomic characteristics, and lifestyle variables in the population of the province of Nova Scotia aged 12 years and older using data derived from the Canadian Community Health Survey, 2007-2008.

1.3 Thesis Overview

The rest of the thesis is organized as follows. Chapter 2 provides a concise literature review on insomnia for the years 2000-2010. Chapter 3 describes the data sample, the logistic regression analysis, and the bootstrapping method. Results are reported in Chapter 4. Chapter 5 presents a discussion of the results, conclusions, limitations, and suggestions for future work.

CHAPTER 2 – LITERATURE REVIEW

In this chapter, an overview of definitions of insomnia and its prevalence and epidemiology in the populations of different countries, and in Canada in particular, is presented. In addition, factors which are expected to be risk factors for developing insomnia are discussed. Factors associated with an elevated occurrence of a disease are generally termed *risk factors* (Taylor, Lichstein, & Durrence, 2003). This literature review covers articles published after the year 2000. However, the studies considered here also refer to other research articles published before the year 2000.

2.1 Defining Insomnia

There has been and continues to be a number of definitions to assess insomnia. Sutton, Moldofsky, and Badley (2001) define insomnia as difficulty initiating or maintaining sleep. Kim et al. (2001) uses a similar definition: difficulty initiating and maintaining sleep, and early morning awakening. Liu et al. (2000) uses early morning awakening, medication use, excessive daytime sleepiness, and sleep duration in addition to difficulty initiating or maintaining sleep. The following dimensions were employed to estimate the nature, severity, and impact of sleep problems by LeBlanc et al. (2007): “severity of sleep onset, sleep maintenance, and early morning awakening problems; sleep satisfaction; interference of sleep difficulties with daytime functioning; noticeability of sleep problems by others; and distress caused by the sleep difficulties.” Ohayon and Bader (2010) utilized a six-point frequency

scale to assess insomnia symptoms as difficulty initiating sleep, difficulty maintaining sleep, and non-restorative sleep.

While the aforementioned studies employ similar definitions of insomnia, no two definitions are exactly the same. Duration of sleep problems taking into consideration varies between different studies as well. Ohayon, Roberts, Zulley, Smirne, and Priest (2000); Roberts, Roberts, and Duong (2008); and Su, Huang, and Chou (2004) utilize DSM-IV criteria (American Psychiatric Association, 1994). See Appendix A for the definition. Others employ 6-month scales (Gellis et al., 2005; Riedel, Durrence, Lichstein, Taylor, & Bush, 2004; Taylor et al., 2007). One-month or four-week insomnia is used as a dependent variable in multivariate techniques in studies conducted by Katz and McHorney (1998), Kim et al. (2001), and Liu et al. (2000). Here, unlike the common 6-week criteria to define chronic insomnia, authors used a 4-week duration to detect chronic insomnia. Patten et al. (2000) studied sleep problems during the past 12 months.

Mai and Buysse (2008) define insomnia as a disorder rather than a symptom. They report that insomnia is a disorder resulting from “hyperarousal”, i.e. a higher state of alertness. Also, insomnia is viewed as a disorder in a study conducted by Roberts et al. (2008).

Furthermore, Taylor et al. (2003) emphasize the importance of making a distinction between primary insomnia and secondary insomnia explaining that there is a risk of making misleading conclusions about actual causes of insomnia. The majority of research on insomnia has concentrated on primary insomnia (e.g., caused by mood disturbance and arousal at bedtime). However, 70% of all insomnia is secondary insomnia (e.g., caused by medical disorders, such as cancer and pulmonary disease, or psychiatric disorders, such as

mood and anxiety disorders), but there has been little research conducted to examine it (Lichstein, Durrence, Riedel, & Bayen, 2001).

Depending on how insomnia is defined in a particular study, the prevalence rates may vary. For example, the prevalence rate of one-month insomnia in the population of Taiwan aged 65 and over was 6% (Su et al., 2004), while 21.4% of Japanese adults aged 20 and over complained of at least one insomnia symptom (Liu et al., 2000). For Ontario patients with cancer, this rate was 31%, which is higher than the abovementioned prevalence rates (Davidson, MacLean, Brundage, & Schulze, 2002). The lack of a clear definition of insomnia and distinctive ways to assess it makes comparing results between any two studies an almost intractable task.

2.2 Insomnia in the Context of Chronicity

Another aspect of insomnia-related research involves the chronicity of either insomnia or health conditions. Studies in this area may be classified into two categories. The first category includes research on chronic insomnia; the second category includes insomnia in relation to chronic illnesses.

Taylor et al. (2007) conducted a study on the coexistence of chronic insomnia with other medical disorders among 772 subjects aged 20-98 years in Shelby County, Tennessee. They note that people who reported chronic insomnia had been diagnosed with the following medical conditions: heart disease (21.9%), high blood pressure (43.1%), neurologic disorder (7.3%), breathing problems (24.8%), and gastrointestinal problems (33.6%). These proportions are much higher than reported percentages from subjects not affected by chronic insomnia. On the other hand, subjects suffering from heart disease (44.1%), cancer (41.4%), high blood

pressure (44%), neurologic disorder (66.7%), breathing problems (59.6%) and gastrointestinal problems (55.4%) reported suffering from chronic insomnia more of the time than those subjects not diagnosed with the aforementioned medical conditions. All in all, the study concludes that there exists a bidirectional relationship between chronic insomnia and other medical problems, and the treatment of insomnia in the context of such comorbid disorders is yet to be investigated.

Johnson, Roth, Schultz, and Breslau (2006) used a US sample of 1014 early adolescents (13 to 16 years of age) to assess the prevalence and chronicity of insomnia per the DSM-IV definition and criteria. The authors found that insomnia is common among early adolescents at a rate of 10.7%, and reveals chronicity at a median age of 11 years. In particular, 68.5% of the subjects revealed difficulty initiating sleep, 26.2% had trouble maintaining sleep, and 48.1% reported nonrestorative sleep. Moreover, 52.8% of the subjects reported insomnia in the context of comorbid psychiatric disorders. Insomnia was found to be prevalent in 9.4% of the early adolescents with or without comorbid disorders. In terms of gender differences, the study found that girls were more prone to insomnia than boys.

To determine the association between insomnia and chronic medical illnesses, Katz and McHorney (1998) carried out a cross-sectional analysis of data from 3445 subjects from Medical Outcomes Study (MOS), conducted in 3 cities: Boston, MA; Chicago, IL; and Los Angeles, CA. They defined insomnia as the complaint of initiating or maintaining sleep. The aim of the study was to estimate the prevalence of, and to identify the risk factors for, insomnia. Insomnia was severe and mild in 16% and 34% of the patients, respectively. At a two-year follow-up, the percentages of patients who still had sleep problems were 83% and 59% with severe and mild insomnia, respectively. The following odds ratios related to mild and severe insomnia were reported: current depressive disorder, 2.6 and 8.2; subthreshold depression, 2.2

and 3.4; congestive heart failure, 1.6 and 2.5; obstructive airway disease, 1.6 and 1.5; back problems, 1.4 and 1.5; hip impairment, 2.2 and 2.7; and prostate problems, 1.6 and 1.4. The authors concluded that clinicians should identify disorders that have an effect on insomnia. Moreover, sleep-affecting medical conditions, such as cardiopulmonary disease, painful musculoskeletal conditions, and prostate problems deserve special attention.

Katz and McHorney (2002) conducted another study utilizing MOS in order to establish the association between insomnia and health-related quality of life (HRQOL) when chronic conditions are present in the patients. In this study, insomnia was shown to have a negative effect on health-related quality of life. That is, the level of the decrease in HRQOL for severe insomnia can be compared to that of chronic conditions, such as depression and congestive heart failure.

Power, Perruccio, and Badley (2005) analyzed data from a population-based 2000/2001 Canadian Community Health Survey of 118,336 adults aged 18 and over. The aim of the study was to investigate how arthritis was related to two sleep problems, insomnia and unrefreshing sleep, and to examine the mediating role of pain in these relationships controlling for other chronic medical conditions. Four logistic regression models were used to estimate the association between arthritis and insomnia symptoms and unrefreshing sleep. Model 1 included arthritis only; model 2 included model 1 as well as sociodemographic characteristics, lifestyle factors, and other chronic conditions; model 3 included model 2 with the addition of mental health variables (stress, depression); and model 4 was obtained by inclusion the pain variable to model 3. The prevalence rate of insomnia symptoms in people with arthritis was 24.8%, and 11.9% of unrefreshing sleep. Models 3 and 4 were compared in order to measure the effect of pain as a mediator. The addition of the pain variable in the model 4 decreased the effect of arthritis by 53% for insomnia symptom and by 64% for unrefreshing sleep, indicating

that pain is a mediator in a relationship between arthritis and insomnia. Moreover, all the chronic conditions showed significant associations with insomnia symptoms and unrefreshing sleep, except neurological conditions (multiple sclerosis, or epilepsy), with migraine having the greatest odds ratio.

2.3 Consequences of Insomnia

The studies above focused on factors that caused insomnia. The main aim of some other studies, however, was to examine the consequences of insomnia on people's quality of life and functioning, psychological functioning in particular. Increased attention has been given in the past few decades to the impacts of insomnia on health and functioning (Taylor et al., 2003). On the one hand, insomnia may help to initiate the development of psychological disorders (e.g., depression) acting as a stressor. On the other hand, insomnia may be a symptom of many disorders, such as depression (Taylor et al., 2003).

Ancoli-Israel (2006) presented a review of the association between chronic insomnia and chronic physiological or psychiatric disorders, such as arthritis, depression, heart disease, diabetes, etc. According to the findings, chronic insomnia is most severe and prevalent when it occurs in subjects already suffering from some chronic illness, but this does not exclude a bidirectional relationship. From such a perspective, the study illustrates two categories of the comorbidity of insomnia with other chronic illnesses, mainly for treatment purposes. The first category concerns insomnia complaints associated with a reported chronic illness symptom, such as congestive heart failure, Cheyne-Stokes respiration, or nighttime gastroesophageal reflux disorders. In this group, treatment of insomnia is unlikely to reduce symptoms of related chronic illnesses. The second category includes the coexistence of

insomnia with physiological disorders, such as diabetes and arthritis. In this case, insomnia is viewed as a feature of the chronic disorder, and treatment of the latter may improve sleep.

Furthermore, chronic insomnia was investigated by Roberts et al. (2008) in terms of its consequences for health and functioning of adolescents, aged 11 to 17 years, residing in the Houston metropolitan area, Texas. This was a prospective study performed over a 12-month period using the DSM-IV diagnostic criteria for insomnia. Based on the results, over 25% of adolescents experienced chronic insomnia. This is viewed as a chronic condition occurring for 1 to 4 years in the given age group. The effects of chronic insomnia on interpersonal functioning, somatic and mental health were examined. The study reported that the impact of insomnia on the three aforementioned domains was severe. Moreover, the consequences of chronic insomnia in adolescents can be compared to those of other psychiatric disorders, such as mood or anxiety disorders and psychoactive substance abuse. Also, the authors note that the relation between insomnia and psychological risk factors may be reciprocal being the cause for insomnia and a consequence of it.

Riedel and Lichstein (2000) provide a comprehensive overview of results related to the study of insomniacs and their daytime functioning. In general, evidence in the literature suggests that people suffering from insomnia are not significantly affected by daytime dysfunction. One reason for this seemingly counterintuitive conclusion might be that people with insomnia tend to overstate daytime impairment, probably because of depressive and anxiety symptoms associated with sleep disorders. Moreover, insomnia tends to produce fatigue and mood changes sufficient to affect the quality of life, but it is not reported to cause psychomotor or cognitive dysfunctions, as is well-documented in the literature. The authors report that there exist studies in the literature that have raised doubts about whether poor sleep causes daytime deficits. That is, lack of daytime sleepiness may not be a sign of nighttime

sleep deprivation in people who complain of insomnia. Thus, suggesting that an arousal problem in people who complain of insomnia is accountable for both insomnia and daytime dysfunction, the authors point out to two theories. The first is the hyperarousal theory, according to which people reporting insomnia experience chronic physiological arousal. The second theory concerns chronic psychological arousal, according to which increased anxiety and worry about nighttime sleep could interfere with daytime functioning. All in all, this review paper suggests that daytime deficits in people suffering from insomnia need to be proven.

2.4 Factors Linked to Insomnia

In addition to chronic illnesses, a number of factors have been linked to insomnia. Among those, depressive symptoms, cigarette smoking, alcohol use, socioeconomic status, and levels of physical activity have been investigated broadly.

Patten et al. (2000) employed a US longitudinal study of 7,960 adolescents aged 12 to 18 to examine factors related to the development and persistence of sleep problems. With the utilization of logistic regression analyses, it was found that 28% of sleep problems and 9% of frequent sleep problems were developed by year 1993 (follow-up) in 4,866 adolescents who did not have sleep problems in 1989 (baseline). Those with sleep problems at baseline ($n = 3,094$), reported sleep problems and frequent sleep problems in 52% and 21% cases respectively. Female gender and notable depressive symptoms were significant predictors of sleep problems and frequent sleep problems. The smallest proportion of frequent sleep problems was reported by those who remained being non-smokers from baseline to follow-up. Compared to this category of non-smokers, that proportion almost doubled for adolescents who were non-

smokers at baseline, but became smokers by follow-up. Those who were smokers in 1989, but quit by 1993, had lower proportions of frequent sleep problems in contrast to non-smokers both at baseline and follow-up. As a conclusion, the most important findings of this study were that depressive symptoms and cigarette smoking were significant predictors for the development and persistence of frequent sleep problems in the adolescent population of the United States.

Another study was conducted in Shelby County, Tennessee, by Riedel et al. (2004) to examine the relationship between smoking and chronic insomnia, defined in the study as the problem persisting for 6 months. Sleep was measured through 2-week sleep diaries. When controlled for demographics, health, behavioural, and psychological variables, light smoking (< 15 cigarettes a day) was significant (OR = 2.75) predictor of chronic insomnia, while heavier smoking was not. Light smokers did not have a significant difference from non-smokers on sleep-onset latency, number of awakenings during the night, wake time after sleep onset, or sleep efficiency.

Stein and Friedman (2005) review facts about an association between sleep disorders and insomnia with alcohol use. The study employs a database going back through approximately 40 years of published literature. Collected articles are categorized into behavioural studies and clinical studies. Based on the former, regular use of alcohol for three or more days diminishes the promoting of sleep, thus bringing about sleep disorders and insomnia. Clinical studies collectively suggest that while extensive alcohol use positively affects insomnia, the strength of such association is yet to be determined. One central conclusion of this study is the need for enhanced investigations of alcohol use in patients with insomnia.

Vinson et al. (2010) did a cross-sectional study on 1,984 primary care patients aged 50 and over in the US to explore the relationship between drinking status (none, moderate, or

hazardous) and sleep problems. The results of multivariate analyses indicated no associations between drinking status and any measures of insomnia, overall sleep quality, or restless legs syndrome symptoms. Moderate and hazardous drinking showed 40% reduction in the odds of sleep apnea compared to non-drinkers. Men in the older age group were at risk of developing sleep apnea. A strong association between self-reported use of alcohol for sleep and hazardous drinking was found (OR = 4.5, compared to moderate drinking).

Davidson et al. (2002) conducted a cross-sectional study of sleep disturbance in cancer patients attending six diagnostic groups (clinics) at the Kingston Regional Cancer Centre, Ontario, Canada. This study examined: (a) the prevalence of sleep problems; (b) sleep problem prevalence relative to cancer treatment; and (c) the nature of reported insomnia (type, duration, and associated factors). Excessive fatigue (44% of the patients), leg restlessness (41%), insomnia (31%), and excessive sleepiness (28%) were mostly common. Lung cancer patients were at the highest or second-highest level of prevalence of all sleep problems. Patients with breast cancer had a high prevalence rate of fatigue and insomnia. The relationship between recent cancer treatment and several sleep problems was observed: the prevalence rate of excessive sleepiness and fatigue was significantly higher for patients who reported recent treatment. The most general type of insomnia reported was waking several times (76% of patients). The longest duration (more than 6 months) was reported in 75% of patients. The most common attributions for insomnia were thoughts, pain or discomfort, and concern about their health. The authors conclude that patients with breast and lung cancer are particularly inclined to sleep problems; therefore, special consideration should be given to satisfy their needs.

Gellis et al. (2005) investigated the relationship between insomnia and socioeconomic status (SES) in Shelby County, Tennessee. SES was measured in the number of years by three

levels of education: individual, household, and community. Insomnia was measured as a complaint of difficulty sleeping over a period of at least 6 months, in addition to a sleep pattern described by sleep latency or wake time during the night. According to their findings, the likelihood of insomnia decreased by 13% with each added year of education. This led the authors to the conclusion that education status, both individual and household, is a risk factor for insomnia, even after controlling for age, gender, and ethnicity. The community indicator, however, did not have a relationship with insomnia. In addition, people with lower levels of individual education and those who have insomnia are highly subjected to impairment, defined in the study as fatigue, psychological distress, and daytime sleepiness.

Morgan (2003) explored a relationship between sleep quality and physical and social activity levels in older age groups. The Nottingham Longitudinal Study of Activity and Ageing (NLSAA), an ongoing study among people aged 65 and older, living in the UK, was employed for detailed assessment at three survey waves (1985, 1989, and 1993). The initial survey was carried out in 1985, when 1,042 elderly people were interviewed to measure physical and social activities, mental and physical health status, and sleep quality. Two follow-up reassessments were conducted in 1989 and 1993 on the samples of 690 and 410 elders, respectively. Logistic regression models, adjusted for age, sex, and health status, were applied with insomnia prevalence in three separate waves (1985, 1989, and 1993) as a dependent variable. The results showed that lower physical health, depressed mood, and lower physical activity level were significant risk factors for insomnia among elderly. These findings also suggest that higher levels of repeated daily activities related to recreation and personal and domestic maintenance may protect against late-life insomnia.

2.5 Insomnia in Different Countries

A number of similar studies were also conducted in several European countries, Japan, and Taiwan to assess prevalence of insomnia and its associated factors.

A study by Ohayon et al. (2000) employed a sample of more than 1,100 adolescents aged 15-18 years old residing in France (30%), the UK (28.5%), Germany (21.1%), and Italy (21%) in order to observe sleep disorders per DSM-IV. To provide a reference basis, this adolescent sample was contrasted with data coming from a group of over 2,100 young adults aged 19-24 years old. The collected data involved sleep/wake schedule, insomnia, and sleep habits. Results revealed several differences and similarities between adolescents and young adults. First, 4% of the individuals in each group complied with the criteria of insomnia per DSM-IV. For the adolescents diagnosed with insomnia, sleep disorders were found to be related to the use of psychoactive drugs or some mental disorder. Second, 26% of the individuals in each group reported at least one sleep disorder symptom. Third, approximately 75% of both adolescents and young adults who were diagnosed with anxiety reported at least one symptom of insomnia. Next, in terms of sleep/wake schedules, the authors reported no higher rate of circadian cycle disorders in adolescents than in young adults. Moreover, sleep habits changed drastically between the two groups: sleep among adolescents lasted longer and was associated with fewer sleep disruptions, unlike the case for young adults.

Dregan and Armstrong (2009) conducted a cross-cohort analysis on two longitudinal datasets in the UK: Health and Activity Lifestyle Survey (HALS) 1984-1985 ($n = 7,785$) and English Longitudinal Study of Ageing (ELSA) ($n = 21,834$). Age, period (time effect, such as a period of social crisis) and cohort effects in sleep loss through worry were investigated for people aged 50 and over. The authors found that both age and period affect insomnia. The

prevalence of sleep loss through worry declined with age, specifically between the 50s and late 70s. As for the period effect, this study reports that while a rise in sleep disorders has been prevalent since the beginning of the 20th century, different age patterns at different times and places are likely to yield different results on the effect on insomnia. All in all, the presence of inconsistent or discrepant reports in the literature indicates that measuring insomnia through age per se as a risk factor may not be sufficient without considering and exploring, among other elements, social and temporal factors.

Hajak (2001) conducted research on the prevalence and quality-of-life impact of severe insomnia among 1,913 adults aged 18 years old and over in Germany. Based on collected data involving quality of life, the study reports a 4% prevalence rate of severe insomnia, which was most frequent among women, unemployed subjects, and lonely subjects, but less so among subjects aged 65 years and over. Of the sampled subjects, 74% revealed experiencing severe sleep problems. Moreover, subjects suffering from severe insomnia rated their quality of life as bad 22% of the time and as good 28% of the time. These ratings were contrasted with those subjects that reported no sleep disorders: 3% of the them rated their quality of life as bad, whereas 68% gave a good rating. Per these findings, the author concluded that insomnia (severe or not) was common and chronic in Germany with relevant consequences on the quality of life.

Ohayon and Smirne (2002) carried out a similar study to evaluate the prevalence and consequences of insomnia in Italy. Their sample consisted of 3,970 individuals aged 15 years and over. Results revealed that 27.6% of the subjects reported insomnia symptoms, even though these subjects did not suffer from any mental or physical disease. Moreover, 10% of the subjects reported dissatisfaction with sleep, and 7% of the subjects had insomnia disorder diagnoses. In terms of the prevalence of insomnia in different genders, the study

reported that such symptoms were found in 36.8% of the men and 44.6% of the women in the sample. A logistic regression model was used to reveal that gender, marital status, heart disease, and level of daily stress, to name a few, were not significantly associated with sleep dissatisfaction. The age group of those between 45 and 64 years ($OR = 2.3$) and age group of 65 years and older ($OR = 3.0$); having a depressive disorder ($OR = 2.0$) or an anxiety disorder ($OR = 2.4$); health perceived as being poor ($OR = 2.0$); and having one ($OR = 3.1$), two ($OR = 8.5$), or three or four ($OR = 14.8$) insomnia symptoms were significant predictors for sleep dissatisfaction. Finally, the study reports that dissatisfaction with sleep has significant association with daytime sleepiness and road accidents in middle-aged subjects.

Ohayon and Bader (2010) evaluated the prevalence of insomnia and factors associated with insomnia symptoms in Sweden. The general population between 19 and 75 years old was employed for this study. The proportions for difficulty initiating sleep (DIS), difficulty maintaining sleep (DMS), and non-restorative sleep (NRS) were the following: 6.3%, 14.5%, and 18% respectively; the overall rate of those who had at least one of the aforesaid insomnia symptoms was found to be 32.1%. Restless leg symptoms, depressive mood together with anxious mood, and breathing pauses during sleep were predictors for DIS and DMS, but not for NRS. Living in a large city ($OR = 2.0$) and drinking alcohol daily ($OR = 4.6$), however, were predictors for NRS. Also, there was an association found between DIS, DMS, and NRS and daytime mental or physical fatigue. Therefore, the study concluded that approximately a third of the Swedish population experiences insomnia symptoms at least 4 nights a week with the higher prevalence in women than in men.

Liu et al. (2000) carried out an epidemiological study on insomnia and daytime sleepiness among 3,030 adult subjects in Japan aged 20 years and over. The authors examined the prevalence of sleep disorders in the adult Japanese population as well as the effect of certain

sociodemographic variables, such as age, gender, employment, marital status, etc., and insomnia symptoms on daytime sleepiness. Results reveal that 21.4% of the subjects complained of at least one insomnia symptom. Moreover, middle-aged and elderly subjects report longer sleep duration than younger subjects, despite other studies concluding that the latter need 8-9 hours of sleep. Insomnia, however, was observed to be more widespread among the elderly subjects in the study. Overall, the study concludes that short sleep duration is the most significant predictor of daytime sleepiness in the adult Japanese population.

Another cross-sectional study on 4,000 adults aged 20 years and older was conducted in Japan by Kim et al. (2001). The aim of their study was to estimate the prevalence of somatic and psychological complaints (SPCs) and to examine their association with insomnia in the general adult population of Japan. The overall prevalence rate of SPC was found to be 78.6% with the most prevalent somatic complaints for stiff neck/shoulder (45.3%), backache (35.1%), and fatigue (31.4%). Insomnia was prevalent in 24.1% of those with SPC and 10.9% in those without SPC. Logistic regression analyses showed a strong association of insomnia with the following SPCs after controlling for sociodemographic factors: backache (OR = 1.4), epigastric discomfort (OR = 1.7), headache (OR = 1.7), fatigue (OR = 1.7), irritability (OR = 1.4), and loss of interest (OR = 1.8).

Prevalence and risk factors for insomnia in the Taiwanese population of 65 years and above was assessed by Su et al. (2004). With an overall rate of 6%, 8% of women and 4.5% of men reported one-month insomnia. Nocturnal micturation (OR = 20.6), use of hypnotics (OR = 3.2), body pain (OR = 2.2), and depressive symptoms (OR = 1.9) were strong predictors of insomnia for both genders. Risk factors for insomnia in men were mental disease (OR = 8.6), single marital status (OR = 2.3), and pulmonary disease (OR = 2.7). Women had depressive

symptoms (OR = 2.2), body pain (OR = 2.6), and lack of education (OR = 1.8) as predictors for one-month insomnia.

Insomnia is widespread among the Canadian population as well. Sutton et al. (2001) used data obtained from the Canadian General Social Survey, Cycle 6, (GSS6) that was performed by Statistics Canada in 1991. The aim of the study was to estimate the prevalence of insomnia and to determine factors associated with insomnia in the Canadian population aged 15 and older. They found that insomnia was highly prevalent, reaching a rate of 24%. Results of multivariate logistic regression led to the conclusion that female gender, being widowed or single, low education, low income, not being in the labour force, ever having smoked, life stress, specific chronic physical health problems (circulatory, digestive and respiratory disease, migraine, allergy and rheumatic disorders), pain, activity limitation, and health dissatisfaction were associated with insomnia. Age, however, did not have a significant association with insomnia which brought the authors to the conclusion that insomnia cannot be considered as part of the ageing process.

LeBlanc et al. (2007) conducted a study in the province of Quebec, Canada. Nine hundred fifty three participants were included in the sample. They were categorized in three groups: (1) insomnia syndrome (n = 147), (2) insomnia symptoms (n = 308), and (3) good sleepers (n = 493). According to the findings, insomnia was associated with stress, higher predisposition to arousal, and increased impairment to health quality.

2.6 Conclusions

This literature review presented the definitions of insomnia employed in various studies, the prevalence of sleep problems, and factors associated with sleep problems in different

populations from different parts of the world. Three types of studies were reviewed in this chapter:

- (i) cross-sectional, comparing the prevalence of problems in people with insomnia and without insomnia (Davidson et al., 2002; Katz & McHorney, 2002; Ohayon & Bader, 2010; Power et al., 2005; Sutton et al., 2001; Taylor et al., 2007);
- (ii) prospective, comparing patients' characteristics at the baseline and then at the follow-up (Katz & McHorney, 1998; Morgan, 2003; Patten et al., 2000; Roberts et al., 2008);
- (iii) review reports (Ancoli-Israel, 2006; Ohayon & Smirne, 2002; Stein & Friedmann, 2005; Taylor et al., 2003), including one meta-analysis (Zhang & Wing, 2006).

As was demonstrated in this chapter, sleep problems were described in the context of a number of explanatory variables (sociodemographic and socioeconomic characteristics, lifestyle and psychological factors, and health conditions). In that respect, three distinguished studies assessed the contribution of all these factors to the complex issue of insomnia. Sutton et al. (2001) focused on the impact of the whole spectrum of the above-listed factors, in addition to pain, on insomnia. Despite the common belief that insomnia increases with an increase in age, in their study age did not show a significant association with insomnia. Taylor et al. (2007), on the other hand, found that advanced age (i.e., > 65 old) was highly associated with insomnia. A number of medical problems, such as high blood pressure, heart disease, cancer, asthma, emphysema, diabetes, back pain, migraine, stomach ulcers, irritable bowel syndrome, etc. were included in the study, controlling for demographics, depression and anxiety variables. Power et al. (2005) examined sociodemographic characteristics, including education and income, lifestyle factors, chronic medical conditions as those utilized by Taylor

et al. (2007) with addition of other chronic conditions and mental health, as depression and level of stress.

The majority of the articles summarized in this chapter utilized logistic regression models, which appear to provide an appropriate reference to employ the logistic regression model in this work. In the context of the Canadian population, the following variables were hypothesised to have a relationship with insomnia: age, gender and marital status were included as demographic characteristics; the highest level of household education and the total household income were used as socioeconomic characteristics; alcohol consumption, smoking status, and physical activity were considered as lifestyle factors; and among health conditions, asthma, chronic bronchitis, emphysema, arthritis, high blood pressure, heart disease, diabetes, intestinal or stomach ulcers, bowel disorder, migraine, cancer, and back problems were included.

CHAPTER 3 – METHODOLOGY

3.1 Source of the Data

This thesis uses data from the Canadian Community Health Survey conducted by Statistics Canada in 2007-2008. The Canadian Community Health Survey (CCHS) is a cross-sectional survey that processes information related to health issues for the Canadian population. Data are collected on an ongoing basis and occurs every year, instead of every two years as was the case prior to 2007.

In addition to sociodemographic and administrative data, the core content, the theme content, the optional content, and the rapid response content are included in CCHS. The core content is assembled from all survey respondents. The theme content covers questions concerning a specific topic. The optional content addresses the provincial or regional public health concerns and gets individually selected by health regions. The rapid response component is available for organizations interested in national estimates on any topic related to the health of the population.

3.2 Target Population

The target population consisted of the persons aged 12 years and older residing in private dwellings in Canada. Individuals residing on Indian Reserves or Crown lands, those residing in institutions, full-time members of the Canadian Forces, and residents of certain remote regions were excluded from this survey (Statistics Canada, 2009). The CCHS covered approximately 98% of the Canadian population aged 12 years and older.

3.3 Design of CCHS 2007-2008

Each province was divided into health regions (HR) and each territory was designated as single HR. For CCHS 2007-2008 the data were collected for 121 health regions. The province of interest, Nova Scotia, was divided into 6 health regions. The response rate for the province was 80.5%. To provide reliable estimates for each HR, a CCHS sample size, which was determined to be approximately 130,000 respondents over a period of 2 years, was allocated in three steps. A minimum size of 500 respondents for each HR was imposed in the first step, resulting in total of 60,350 allocated units. In the second step, the rest of the available sample was allocated among the provinces proportional to their population size. The total targeted sample size for Nova Scotia was 5,040. However, this number was increased before collecting data. Taking into consideration the possible non-response rate, the needed sample size was determined to be 7,533. The increased sample size was then adjusted back to the desired number. Finally, in the third step, the provincial sample was allocated among its HRs proportionally to the square root of the estimated population in each HR.

The CCHS uses three sampling frames to select the sample of households: 49% of the sampled households come from an area frame, 50% comes from a list frame of telephone numbers and the remaining 1% comes from a Random Digit Dialing (RDD) telephone number frame. For most of the health regions, 50% of the sample is selected from the area frame and 50% from the list frame of telephone numbers. In two health regions (Nord-du-Québec and Prairie North), only the RDD frame is used. In Nunavut, only the area frame is used. In the Yukon and Northwest Territories, most of the sample comes from the area frame, but a small RDD sample is also selected in the territorial capitals.

Between January 2007 and December 2008, a total of 131,959 valid interviews were conducted using computer assisted interviewing (CAI). Nearly half of the interviews was conducted in person using computer assisted personal interviewing (CAPI), and the other half was conducted over the phone using computer assisted telephone interviewing (CATI). In cases when the selected respondent was incapable of completing an interview for reasons of physical or mental health, another knowledgeable member of the household supplied information about the selected respondent. This is referred to as a proxy interview. While individuals interviewed this way were able to provide accurate answers to most of the survey questions, the more sensitive or personal questions were usually beyond the scope of knowledge of a proxy respondent. This resulted in some questions from the proxy interview remaining unanswered. Proxy interviews were attempted to be kept to a minimum.

3.4 Statistical Analyses

All statistical analyses were performed using SAS software version 9.2. Estimates of the prevalence of insomnia in the sample of 5,018 persons residing in Nova Scotia were calculated. The variables anticipated to be related to insomnia were selected based on the previous studies. After a thorough examination of the variables selected from the data sample and with an application of stepwise logistic regression procedure, the most important variables to be included in the final model were determined. Bivariate logistic regression analyses were carried out to calculate odds ratios for each of the explanatory variables in the final model. Then, multivariate logistic regression analyses were performed with all the variables entered at once. Odds ratios were computed for each variable, controlling for all other explanatory variables, to estimate the risk of having insomnia.

The intent of this study was to assess the association between insomnia and health factors, demographic and socioeconomic characteristics, and lifestyle variables. Therefore, mental health variables such as depression, anxiety, and stress were not included in the study even though the association between them and insomnia has been determined in the literature (Koffel & Watson, 2009; LeBlanc et al., 2007).

3.5 Data Availability

Since the subject of sleep problems is an optional content in the Canadian Community Health Survey, the only available data for the questions concerning insomnia were for the province of Nova Scotia. This data availability set a limit on the sample size, which equalled to 5,152. Since 134 cases had missing values for the variable for sleep problems, statistical analyses were performed on a data set containing 5,018 cases.

3.6 Response Variable

Different questionnaires and definitions of insomnia have been used in a variety of studies as summarized in Chapter 2. For this study, the subjects were asked the following question: “How often do you have trouble going to sleep or staying asleep?” The response options were “none of the time”, “a little of the time”, “some of the time”, “most of the time”, or “all of the time.” In this study, responses of “most of the time” and “all of the time” were grouped to represent the presence of insomnia, while responses of “none of the time”, “little of the time”, and “some of the time” were considered to represent the absence of insomnia. Morgan (2003) used a similar classification in his study of relationships between

sleep quality and physical and social activity levels in older age groups. Therefore, the response variable INSOMNIA was obtained from the Statistics Canada derived variable slp_02. If slp_02 was in categories 1, 2, or 3, then INSOMNIA equalled 0, denoting absence of sleep problems. If slp_02 was in categories 4 or 5, then INSOMNIA equalled 1, denoting presence of sleep problems. See Appendix A for the categories of the variable slp_02.

3.7 Explanatory Variables

A number of possible determinants of sleep difficulties have been specified in the literature. Demographic and socioeconomic characteristics, lifestyle variables, and health factors have all been connected to sleep problems. Therefore, they were included in the multivariate analyses carried out in this work.

Demographic characteristics included age, gender, and marital status. The variable for age was grouped in 7 categories: 12-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75 and older. The Statistics Canada variable for gender was unchanged, and the Statistics Canada derived variable for marital status was recoded to yield three codes representing those who were married or living in common law; those who were widowed, separated, or divorced; and those who were single or have never been married.

The highest level of household education and the total household income were used as socioeconomic characteristics. Education was retained as a four-category variable as defined by Statistics Canada: less than secondary school graduation, secondary school graduation, some post-secondary, and post-secondary graduation. Household income was grouped to represent five categories of household income: less than \$20,000; \$20,000-39,999; \$40,000-59,999; \$60,000-79,999; and \$80,000 or more.

Among lifestyle factors, alcohol consumption, smoking status, and physical activity index were included in the analyses. The Statistics Canada derived variable for type of drinker was retained having three categories denoting regular drinkers, occasional drinkers, and those who have not had a drink over the last 12 months. The Statistics Canada derived variable for type of smoker was recoded to represent current smokers (those who smoke daily, those who smoke occasionally (former daily smokers), and those who have always been occasional smokers); former smokers (those who have been former daily smokers or former occasional smokers); and those who have never smoked. The Statistics Canada derived variable for physical activity index was unchanged and grouped as active, moderately active, and inactive.

In addition to demographic, socioeconomic, and lifestyle variables, a sequence of dichotomous variables was created for the following health conditions: asthma, chronic bronchitis, emphysema, arthritis, high blood pressure, heart disease, diabetes, intestinal or stomach ulcers, bowel disorder, migraine, cancer, and back problems excluding fibromyalgia and arthritis.

There are two types of variable designations in this study. The variables of the first type are written in uppercase letters. These variables are derived from Statistics Canada variables and are used in the analyses performed with SAS. The variables of the second type are written in lowercase letters. These variable labels correspond to the variable names taken from Statistics Canada. The recoding designations for all the variables used in the study are listed in Table 19 of Appendix B.

3.8 Logistic Regression Model

A logistic regression model describes the relationship between a categorical outcome variable, the response, and a set of explanatory variables, the predictors. The response variable is usually dichotomous. Dichotomous responses are those that have two possible outcomes - most often they are “disease” and “no disease”. However, if the response variable has more than two response levels, it is called polytomous. The explanatory variables may be continuous or discrete or a mixture of both. Therefore, the logistic regression model can be thought of as a binomial model for the two possible values of the response variable, such as “failure” or “success”, where one or more explanatory variables are used to explain the probability of success. If the two values of outcome are 0 or 1, the mean is the proportion of 1's.

Using logistic regression models has several advantages. Both hypothesis tests and confidence intervals are computed for the model parameters. Hypothesis tests help to identify which predictors affect the response variable; the size of the estimated parameters and the widths of their confidence intervals help assess the strength and importance of these effects (Friendly, 2000). The logistic regression models have a high degree of flexibility since the continuous explanatory variables do not have to be normally distributed and other variables can be dichotomous. Also, the results are easily interpreted because the model coefficients are transformed into odds ratios. The odds ratio is a measure of the strength of association that shows how the outcome is more likely (or unlikely) to occur among one category than the other (Hosmer & Lemeshow, 2000).

The multiple regression model with q -explanatory variables is written as

$$y = \mu + \varepsilon , \quad (3.1)$$

where ε is the error term and

$$\mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q . \quad (3.2)$$

Here, x_1, x_2, \dots, x_q are the explanatory variables, and $\beta_0, \beta_1, \dots, \beta_q$ are the parameters of the model.

This model is suitable for continuous response variables with, conditional on the values of the explanatory variables, a normal distribution with constant variance. Modeling the expected value of this type of response directly as a linear function of explanatory variables leads to fitted values of the response probability outside the range $[0, 1]$. If the binary response is written as:

$$y = \pi(x_1, x_2, \dots, x_q) + \varepsilon , \quad (3.3)$$

then the assumption of normality for ε is wrong. Here, ε may assume only one of two possible values: if $y = 1$, then $\varepsilon = 1 - \pi(x_1, x_2, \dots, x_q)$ with probability $\pi(x_1, x_2, \dots, x_q)$; if $y = 0$, then $\varepsilon = -\pi(x_1, x_2, \dots, x_q)$ with probability $1 - \pi(x_1, x_2, \dots, x_q)$.

Thus, ε has a distribution with mean zero and variance equal to $\frac{\pi(x_1, x_2, \dots, x_q)}{1 - \pi(x_1, x_2, \dots, x_q)}$. The conditional distribution of binary response in this case follows a binomial distribution with probability given by the conditional mean $\pi(x_1, x_2, \dots, x_q)$.

Instead of modeling the expected value of the response directly as a linear function of explanatory variables, the logistic or *logit* function of π is modeled, which gives the logistic regression model:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q. \quad (3.4)$$

The *logit* of a probability is the natural logarithm of the odds of the response that takes the value 1. Equation (3.4) can be rewritten as

$$\pi(x_1, x_2, \dots, x_q) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q}}. \quad (3.5)$$

The *logit* function can take any real value, but the associated probability always lies in the required $[0, 1]$ interval (Everitt & Hothorn, 2006).

Rather than proportions, odds ratios are employed in logistic regression. Odds ratios are the ratio of the proportions for the two possible outcomes:

$$\text{Odds}(Y = 1) = \frac{\pi}{1 - \pi}. \quad (3.6)$$

3.9 Bootstrap

Introduced by Efron in 1979, “bootstrap is a general-purpose technique for obtaining estimates of the properties of statistical estimators without making assumptions about the distribution giving rise to the data” (Harrell, 2001). Using resampling methods, such as bootstrap, relaxes the requirement for data to be normally distributed, which is the most important condition for reliable conclusions about a population in a large proportion of statistical analyses (Moore & McCabe, 2005). The key idea of the bootstrap is to repeatedly simulate a sample of size n from the cumulative distribution function F , compute the statistic of interest, and then evaluate the behaviour of that statistic over B repetitions (Harrell, 2001).

The bootstrap is a data-based simulation method for making statistical inferences (Efron & Tibshirani, 1998). Inference about the population needs random sampling. Most of the time, taking many random samples from the population can be difficult to obtain. However, many resamples can be drawn with replacement from just one random sample. When sampling with replacement, an observation that was randomly drawn from the original sample, gets placed back after being recorded and before the next observation is drawn. Each resample obtained in this way is a simple random sample of the same size as the original one (Moore & McCabe, 2005).

The method of bootstrapping was initially used as a procedure to obtain an estimate of the standard error of an estimator. Moreover, another important purpose of bootstrap methodology is that it can be used to construct confidence intervals for statistics of interest. This non-parametric technique of bootstrapping can be a useful alternative to inference based on parametric assumptions of the distributions of the parameter estimates (Efron & Tibshirani, 1998).

In general, using more bootstrapped replications produces more ‘replicable’ results that are similar to the previous analyses. These results can be repeated by re-running the bootstrap analysis a large number of times. At least 1,000 repeated samples are needed to get reliable results (Barker, 2005).

The principle of bootstrap can be applied to more complicated data structures such as regression models, and logistic models in particular. In this thesis, the bootstrap analyses were run to assess the accuracy of primary statistical results. Odds ratios for the variables in the final logistic regression model with their confidence intervals were computed using 10,000 bootstrap samples. The empirical distributions of the odds ratios were estimated along with their parameters and are presented in Section 4.3.

3.10 Bootstrap Estimates

The basic idea behind the bootstrap method is to treat the sample as the population and to perform a Monte-Carlo-style procedure on it. This is done by randomly drawing a large number of resamples of size n from the original sample (of size n) with replacement. Since the elements in these resamples vary slightly, the statistics $\hat{\theta}$ calculated from these resamples vary slightly. The main assertion of bootstrap is that the relative frequency distribution of these $\hat{\theta}$ values is an estimate of the sampling distribution of $\hat{\theta}$.

The problem of estimating the location parameter (θ) and its standard deviation is considered by the bootstrap method. The sample data x_1, x_2, \dots, x_n can be viewed as realizations of n independent random variables X_1, X_2, \dots, X_n with common distribution function F . Since $\hat{\theta}$ is a function of the random variables X_1, X_2, \dots, X_n , it has a probability distribution determined by n and F . Two problems may be encountered in attempting to determine this sampling distribution: (i) F is known, but $\hat{\theta}$ is a complicated function of X_1, X_2, \dots, X_n , so that its distribution is beyond one's analytical abilities and (ii) F is not known. In what follows, these two cases are illustrated in more detail:

(i) F is known

A large number of samples, say B , of size n from F is generated; from each sample the value of $\hat{\theta}$, say $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$, is calculated. The empirical distribution of these values is an approximation to the distribution function of $\hat{\theta}$. The standard deviation of $\hat{\theta}$ is approximated by calculating the standard deviation of $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$.

(ii) F is not known

The idea of the nonparametric bootstrap is to simulate data from the empirical distribution function F_n . Here F_n is a discrete probability distribution which assigns a probability $1/n$ to each observed value x_1, x_2, \dots, x_n . A sample of size n from F_n is thus a sample of size n drawn with replacement from the collection x_1, x_2, \dots, x_n . The values $\widehat{\theta}_1^*, \widehat{\theta}_2^*, \dots, \widehat{\theta}_B^*$ are calculated from B bootstrap samples of size n from the collection x_1, x_2, \dots, x_n . The standard deviation of $\widehat{\theta}$ is then estimated by

$$s_{\widehat{\theta}} = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\widehat{\theta}_i^* - \widehat{\theta}^*)^2}, \quad (3.7)$$

where

$$\widehat{\theta}^* = \frac{1}{B} \sum_{i=1}^B \widehat{\theta}_i^*. \quad (3.8)$$

CHAPTER 4 – RESULTS

4.1 Data Screening and Cleaning

The data set for the province of Nova Scotia contained 1,195 variables and 5,152 cases. Of the total number of cases, 134 cases were deleted due to having missing values for the variable of interest, slp_02 (later recoded as INSOMNIA, see Table 19 in Appendix B). Based on the literature review and the available data set, 21 variables of interest were selected for the analyses. Prior to analyses, the variables for trouble sleeping, gender of respondent, age of respondent, marital status, type of drinker, level of education in the household, total household income, physical activity index, cigarette smoking, asthma, chronic bronchitis, emphysema, arthritis, high blood pressure, heart disease, diabetes, intestinal or stomach ulcers, bowel disorder, migraine, cancer, and back problems excluding fibromyalgia and arthritis were examined for accuracy of data entry, within range and missing values, and plausible means and standard deviations.

The response variable is categorical. The predictor variables are categorical (e.g., gender) or ordinal (e.g., age). There are no continuous variables in the data set.

4.1.1 *Missing Data*

The variables were examined separately for the two groups: 981 persons with sleep problems and 4,037 persons without sleep problems. Percentages of missing values for each variable were obtained for separate groups and are presented in Table 1. The recoding designations for all the variables in Table 1 are listed in Table 19, Appendix B.

Table 1. Percentages of missing values

Variable	Missing values (%)	
	No insomnia	Yes insomnia
GENDER	0.00	0.00
AGE	0.00	0.00
MARITAL	0.27	0.10
ALCOHOL	0.42	0.41
EDUCAT	5.99	6.32
INCOME	14.44	12.13
PHYSACT	0.02	0.10
SMOKE	0.10	0.00
ASTHMA	0.00	0.00
BRONCH	27.59	29.36
EMPHYS	21.15	12.95
ARTHRIT	0.00	0.00
BLOODPR	0.27	0.51
HEART	0.25	0.41
DIABETES	0.02	0.00
ULCERS	0.12	0.51
BOWEL	0.07	0.31
MIGRAINE	0.05	0.10
CANCER	0.07	0.20
BACKPROB	0.02	0.31

As can be noticed from Table 1, there were more than 14% of cases with missing values on INCOME in the “no insomnia” group, and 12% of cases with missing values in the “yes insomnia” group. The variable BRONCH had nearly 28% of cases with missing values in the “no insomnia” group, and about 29% cases with missing values in the “yes insomnia” group. For EMPHYS, the percentages of missing values were 21% and 13% in the “no insomnia” and the “yes insomnia” groups, respectively. The responses of “not applicable”, “not stated”, or “don’t know” were recoded as “missing”. In this study, the percentages of missing values for the abovementioned variables were higher than 5%. Therefore, variables INCOME, BRONCH, and EMPHYS were good candidates for deletion from the subsequent analyses if the aforementioned recommendation were to be followed (Tabachnick & Fidell,

2007). However, there exists a better approach to deal with missing data, which is illustrated in the following subsection.

4.1.2 Analyzing Patterns of Missing Data

Considering that non-response itself represents data, it is possible to make it useful. If all the subjects with missing values are put in a separate category, a null hypothesis that non-responders are no different from responders on insomnia can be tested (Howell, 2009). Therefore, the test of the equality of two proportions of missing and non-missing data was conducted in order to make a decision whether we can retain the three variables in further analyses. Since the data are nominal, i.e. there is no assumption about the order of the categories, an appropriate test to use in this case was the χ^2 test of independence. This is a nonparametric test as it is not concerned with the parameters of normal distribution. A χ^2 test statistic is computed and then compared to the critical value denoted as χ^2_{cv} (Hurlburt, 2005).

After examining variable EMPHYS, it was observed that it has the age restriction of 30 and over. Out of 981 total missing cases on this variable, 979 cases came from the age categories below 30 years old. The fact that the data in this work is secondary poses difficulties in examining why much data are missing on this variable. Thus, the responses to questions regarding bronchitis are not verifiable. One reason, however, why data on income might be missing is that sometimes people refuse to reveal this type of personal information.

The χ^2 test of independence was performed to see whether there is a difference between non-missing and missing data for the three variables: INCOME, BRONCH and EMPHYS. The test results are presented in Table 2. The results for INCOME were:

$\chi^2 = 3.50$, $df = 1$, probability $p = .06$, and the results for BRONCH were: $\chi^2 = 1.22$, $df = 1$, $p = .27$. These indicate that the null hypothesis of equal proportions cannot be rejected.

Therefore, the proportions between missing and non-missing data in the population are inferred to be equal; hence, we include INCOME and BRONCH in the subsequent analyses. The small p -value ($< .0001$) for the χ^2 test for the variable EMPHYS indicates that the null hypothesis of equal proportions can be rejected, leading to the conclusion that non-missing and missing data are significantly different. Hence, the variable EMPHYS was deleted from the subsequent analyses, leaving a data set with 19 explanatory variables.

Table 2. The results of χ^2 test of independence

Variable	DF	Chi-square statistic	P
INCOME	1	3.5028	.0613
BRONCH	1	1.2185	.2697
EMPHYS	1	33.8098	< .0001

Missing data are generally classified into three categories: MCAR (missing completely at random), MAR (missing at random), and MNAR (missing not at random). Therefore, the pattern of missing data was analyzed in order to determine which of the three abovementioned categories is implied in this study.

Tabachnick and Fidell (2007) state that “the pattern of missing data is more important than the amount missing”. Schwab (2002) suggests the following procedure to analyze the pattern of missing data. First, the diagnostic variable CNT is created. CNT counts the number of variables with missing data for every case. Next, the dichotomous variable PATTERN is created with valid/missing categories. These two variables are used to check if

there are any cases to be considered for removal from further analyses. It was observed that CNT = 4 for two cases, CNT = 3 for 41 cases, CNT = 2 for 296 cases, CNT = 1 for 1,783 cases, and CNT = 0 for the remaining 2,896 cases. No cases were considered for removal.

Thereafter, the frequency distribution of the pattern variable was examined to see if there were frequently occurring patterns of missing data. Table 3 shows these patterns along with the corresponding frequencies. The largest number of patterns (1,194) was for the pattern containing the variable BRONCH alone. The next largest count of patterns (416) was for the pattern that had INCOME alone. The third largest frequency (132) was for two patterns: INCOME and BRONCH together and EDUCAT as an only variable. The sequence of the variables in the pattern in Table 3 is as follows: GENDER, AGE, ASTHMA, MARITAL, ALCOHOL, EDUCAT, INCOME, PHYSACT, SMOKE, BRONCH, BLOODPR, HEART, ULCERS, BOWEL, ARTHRIT, DIABETES, MIGRAINE, CANCER, and BACKPROB.

The χ^2 test of independence was conducted to establish whether the missing and valid cases are statistically different for all of the other variables in the analysis, showing an overall pattern of relationships. The resulting probability p was equal to .82 indicating that the missing and valid groups are equal. Therefore, the missing cases can be characterized as random.

Since there is no rigid number for the proportion of missing cases required for a variable to be eliminated from the analyses, INCOME and BRONCH were kept for further examination. Using methods for dealing with missing data, such as estimating (or imputing) missing values, is beyond the scope of the present study; thus, all the cases with missing values were dropped from further analyses.

Table 3. Patterns of missing data

Pattern	Frequency	Percent
~~XX~X~~~~~	1	0.02
~~X~X~~~~~	2	0.06
~~X~~~~X~~~~~	3	0.12
~~X~~~~~	6	0.24
~~~XXX~~~~~	10	0.44
~~~X~X~~~~~	4	0.52
~~~X~~~~X~~~~~	1	0.54
~~~X~~~~~	5	0.64
~~~XXX~X~~~~~	1	0.66
~~~~XX~X~X~~~~~	1	0.68
~~~~XX~X~~~~~	24	1.16
~~~~XX~X~~~~~	1	1.18
~~~~XX~~~~X~~~~~	1	1.2
~~~~XX~~~~~	97	3.13
~~~~X~~~~X~X~~~~~	2	3.17
~~~~X~~~~X~~~~~	32	3.81
~~~~X~~~~X~~~~~	1	3.83
~~~~X~~~~~X~~~~~	1	3.85
~~~~X~~~~~X~~~~~	1	3.87
~~~~X~~~~~	132	6.5
~~~~~X~XX~~~~~	1	6.52
~~~~~X~X~~~~~	132	9.15
~~~~~X~~~~X~~~~~	2	9.19
~~~~~X~~~~X~~~~~	3	9.25
~~~~~X~~~~~X~~~~~	2	9.29
~~~~~X~~~~~X~~~~~	1	9.31
~~~~~X~~~~~X~~~~~	2	9.35
~~~~~X~~~~~X~~~~~	1	9.37
~~~~~X~~~~~	416	17.66
~~~~~X~X~~~~~	1	17.68
~~~~~XX~~~~~	1	17.7
~~~~~X~~~~~	2	17.74
~~~~~XX~~~~~	3	17.8
~~~~~X~X~~~~~	3	17.86
~~~~~X~XX~~~~~	1	17.88
~~~~~X~~~~~X~	2	17.92
~~~~~X~~~~~	1194	41.71
~~~~~X~~~~~	9	41.89
~~~~~X~~~~~X~	1	41.91
~~~~~X~~~~~	4	41.99
~~~~~X~~~~~	6	42.11
~~~~~X~~~~~	4	42.19
~~~~~X~~~~~	3	42.25
~~~~~X~~~~~	2	42.29
~~~~~	2896	100

### **4.1.3 *Univariate Outliers***

Dichotomous variables with an uneven split between categories are considered as univariate outliers and should be deleted. Tabachnick and Fidell (2007) suggest “deleting dichotomous variables with 90-10 splits between categories, or more, both because the correlation coefficients between these variables and others are truncated and because the scores for the cases in the small category are more influential than those in the category with numerous cases.” The dichotomous variables ULCERS and CANCER were not evenly split within both groups (“no insomnia” and “yes insomnia”), as shown in Table 4. Consequently, they were deleted from the subsequent analyses. The variables HEART and BOWEL had an approximately marginal 92-8 split and 93-7 split, respectively, in the “no insomnia” group. It was decided to preserve them in subsequent analyses. The variables ASTHMA, DIABETES, and MIGRAINE all had marginal 91-9 split in “no insomnia” group as well, so it was decided to retain them here.

Variable EMPHYS had 99-1 split in the “no insomnia” group and 98-2 split in the “yes insomnia” group. However, this variable was previously removed due to the fact that it had a lot of missing values. An uneven split for BRONCH was 97-3 in the “no insomnia” group and about 92-8 in the “yes insomnia” group. Thus, BRONCH was removed from further analyses.

The variables BRONCH, EMPHYS, ULCERS, and CANCER were excluded from the remaining analyses, thus leaving a sample size of 5,018: 4,037 cases in the “no insomnia” group and 981 cases in the “yes insomnia” group. Sixteen explanatory variables were included in further analyses.

Table 4. Variable split in two different groups

Variable	No insomnia (0)			Yes insomnia (1)		
	0	1	Split	0	1	Split
GENDER	1749	2288	43-57	334	647	34-66
ASTHMA(us)	3687	350	91-9	867	114	88-12
BRONCH (us)	2832	91	97-3	638	55	92-8
EMPHYS (us)	3153	30	99-1	834	20	98-2
ARTHRIT	3011	1026	75-25	592	389	60-40
BLOODPR	3037	989	75-25	671	305	69-31
HEART (us)	3701	326	92-8	867	110	89-11
DIABETES (us)	3691	345	91-9	873	108	89-11
ULCERS (us)	3894	138	97-3	899	77	92-8
BOWEL (us)	3769	265	93-7	835	143	85-15
MIGRAINE (us)	3679	356	91-9	795	185	81-19
CANCER (us)	3939	95	98-2	947	32	97-3
BACKPROB	3149	887	78-22	600	378	61-39

*(us) - A variable has uneven split in at least one insomnia group ("yes" or "no")*

#### 4.1.4 Detecting Multivariate Outliers

Cases with an unusual combination of scores for a number of variables are referred to as multivariate outliers (Tabachnick & Fidell, 2007). While a particular variable value may not be a univariate outlier, it may lead to multivariate outliers when considered with other

variables. Outliers, in general, affect results of statistical analyses. Thus, detecting and removing multivariate outliers is an important step.

Multivariate outliers were sought within each group separately by using two methods: Method 1 uses a technique for detecting outliers described by Abraham and Ledolter (2006) and as follows: if the leverage for a case exceeds twice the average leverage, that is

$$h_{ii} > 2\bar{h} = \frac{2(k+1)}{n} , \quad (4.1)$$

where  $k$  is the number of explanatory variables and  $n$  is the number of cases, then that case is a high-leverage case. For the “yes insomnia” group,

$$2\bar{h} = \frac{2(16+1)}{981} = 0.03466.$$

In total, 36 cases with  $h_{ii} > 0.03466$  were identified.

For the “no insomnia” group,

$$2\bar{h} = \frac{2(16+1)}{4037} = 0.00842.$$

In this group, the number of cases with  $h_{ii} > 0.00842$  equaled to 366.

Method 2 uses another approach for detecting multivariate outliers, which is discussed in Tabachnick and Fidell (2007). The criterion for detecting a multivariate outlier is the Mahalanobis distance that is estimated using the  $\chi^2$  distribution at  $\alpha < .001$ . The critical  $\chi^2$  value for the Mahalanobis distance with degrees of freedom equal to the number of variables (16 in this case) is translated to leverage by using the following formula:

$$h_{ii} = \frac{\text{Mahalanobis distance}}{n - 1} + \frac{1}{n} , \quad (4.2)$$

where  $n$  denotes the number of cases.

The critical value of the Mahalanobis distance with 16 variables at  $\alpha = .001$  is 39.252.

Using the equation above, we converted this to a critical leverage value for each group. For the “yes insomnia” group:

$$h_{ii} = \frac{39.252}{981 - 1} + \frac{1}{981} = 0.04107.$$

Eight cases with  $h_{ii} > 0.04107$  were detected in this category.

For the “no insomnia” group:

$$h_{ii} = \frac{39.252}{4037 - 1} + \frac{1}{4037} = 0.00997.$$

In this group, the number of cases with  $h_{ii} > 0.00997$  was equal to 185.

Three logistic regression analyses were run for the three models and the results are compared in Section 4.2.2:

1. Model 1 - a data set with all cases;
2. Model 2 - a data set without the cases detected by Method 1;
3. Model 3 - a data set without the cases detected by Method 2.

#### **4.1.5 Correlations**

Since the correlations between the response variable and each predictor should be strong, but the correlations between explanatory variables shouldn't be high to avoid



Table 5. Spearman's rank correlation and *p*-values between INSOMNIA and explanatory variables

	INSOMNIA
GENDER	0.075 < .0001
AGE	0.031 .026
MARITAL	-0.027 .053
ALCOHOL	0.012(ns) .382
EDUCAT	0.023(ns) .118
INCOME	0.008(ns) .594
PHYSACT	0.060 < .0001
SMOKE	0.040 .005
ASTHMA	0.040 .004
ARTHRIT	0.125 < .0001
BLOODPR	0.060 < .0001
HEART	0.044 .002
DIABETES	0.034 .016
BOWEL	0.117 < .0001
MIGRAINE	0.129 < .0001
BACKPROB	0.152 < .0001

(ns) - Correlation is non-significant at the .05 level.

Table 6. Spearman's rank correlations between predictor variables

	GENDER	AGE	MARITAL	PHYSACT	SMOKE	ASTHMA	ARTHRITIS
GENDER	1.0	0.061	0.021	0.107	-0.081	0.064	0.094
		.131	.131	< .0001	< .0001	< .0001	< .0001
AGE		1.0	-0.291	0.225	0.236	-0.097	0.424
			< .0001	< .0001	< .0001	< .0001	< .0001
MARITAL			1.0	-0.095	-0.197	0.050	-0.109
				< .0001	< .0001	.0005	< .0001
PHYSACT				1.0	0.043	-0.020	0.159
					.002	.162	< .0001
SMOKE					1.0	-0.033	0.120
						.019	< .0001
ASTHMA						1.0	0.006
							.653
ARTHRIT							1.0

Table 6 (Continued)

	BLOODPR	HEART	DIABETES	BOWEL	MIGRAINE	BACKPROB
GENDER	0.054	-0.033	-0.037	0.111	0.104	0.023
	.0001	.020	.010	< .0001	< .0001	.109
AGE	0.448	0.298	0.243	0.067	-0.115	0.128
	< .0001	< .0001	< .0001	< .0001	< .0001	< .0001
MARITAL	-0.087	-0.044	-0.036	-0.049	0.020	-0.052
	< .0001	.002	.010	.0005	.166	.0002
PHYSACT	0.164	0.098	0.104	0.072	-0.021	0.073
	< .0001	< .0001	< .0001	< .0001	.142	< .0001
SMOKE	0.103	0.088	0.088	0.050	-0.026	0.058
	< .0001	< .0001	< .0001	.0004	.070	< .0001
ASTHMA	-0.018	-0.001	-0.014	0.031	0.071	0.058
	.209	.953	.316	.029	< .0001	< .0001
ARTHRIT	0.285	0.209	0.160	0.149	0.023	0.255
	<0.001	< .0001	<0.001	< .0001	.098	< .0001
BLOODPR	1.0	0.257	0.275	0.059	-0.046	0.107
		< .0001	< .0001	< .0001	.001	< .0001
HEART		1.0	0.190	0.085	-0.023	0.118
			< .0001	< .0001	.104	< .0001
DIABETES			1.0	0.054	-0.027	0.059
				.0001	.059	< .0001
BOWEL				1.0	0.082	0.115
					< .0001	< .0001
MIGRAINE					1.0	0.095
						< .0001
BACKPROB						1.0

multicollinearity (Tabachnick & Fidell, 2007), the correlations between the response variable and the predictor variables were calculated first to see if there are any variables that do not have significant correlations with the response variable. The correlation matrix of all the independent variables was constructed using Spearman's rank correlation. Table 5 provides these correlations. Three variables, namely ALCOHOL, EDUCAT, and INCOME had non-significant *p*-values, indicating no correlation with insomnia; thus, they were deleted from further analyses.

Spearman's rank correlations between predictor variables were computed next and are presented in Table 6. The numbers in italic represent the *p*-values. According to the correlation matrix illustrated in Table 6, there are no high correlations between predictors.

#### **4.1.6 Multicollinearity**

The test for multicollinearity by using the Variance Inflation Factor (VIF) was run on Model 1 (with all the cases), from which ALCOHOL, EDUCAT, and INCOME were dropped, leaving 13 explanatory variables. If  $VIF > 10$ , then multicollinearity is present (Tabachnick & Fidell, 2007). In this study, the maximum VIF was 1.75 for AGE which implies that no multicollinearity is evident. Multicollinearity can also be evaluated by the condition index which is obtained in SAS with Collinearity Diagnostics. "Criteria for multicollinearity are a conditioning index greater than 30 for a given dimension coupled with variance proportions greater than 0.50 for at least two different variables" (Tabachnick & Fidell, 2007). In present case, the maximum condition index was approximately 16, and there was no dimension (row) with more than one variance proportion larger than 0.5. The test for multicollinearity was also performed on two other data sets, Model 2 and Model 3,

which were referred to in Section 4.1.4. No multicollinearity was detected for both these models as well.

## **4.2 Logistic Regression Analysis**

### ***4.2.1 Prevalence of insomnia***

Table 7 shows the prevalence rates of insomnia in the sample of Nova Scotia aged 12 and older in the 2007-2008 CCHS dataset. The dataset consists of 2,083 males (41.5%) and 2,935 females (58.5%). The overall prevalence rate of insomnia is 19.5%. It can be seen from the table that the prevalence rate of insomnia is 16% in males 22% in females. Moreover, the overall prevalence of insomnia is higher in the 45-54 age group (25.7%), followed by the 55-64 age group (23.8%) for both males and females. These results also comply with the results of logistic regression analysis presented in Table 10 in Section 4.2.3.

Table 8 shows prevalence rates for medical problems in both people with insomnia and people without insomnia. People with insomnia have greater prevalence rates for asthma, arthritis, high blood pressure, heart disease, diabetes, bowel disorder, migraine, and back problems.

Table 7. Prevalence rates of insomnia in the sample

<b>Characteristics</b>	<b>N</b>	<b>% of Sample</b>	<b>% with Insomnia</b>
<i>Age</i>			
1=12-24	704	14.0	11.8
2=25-34	564	11.2	17.9
3=35-44	709	14.1	20.3
4=45-54	838	16.7	25.7
5=55-64	902	18.0	23.8
6=65-74	659	13.2	17.8
7=75+	642	12.8	16.5
<i>Gender</i>			
Male	2083	41.5	16.0
Female	2935	58.5	22.0
<i>Cigarette Smoking</i>			
1=never smoked	1661	33.1	15.1
2=current (daily/ occasional/ always occasional)	1138	22.7	26.1
3=former (former daily/ former occasional)	2215	44.1	19.6

Table 8. Prevalence of medical problems

<b>Medical Problem</b>	<b>Prevalence of Medical Problem, %</b>	
	<b>People with Insomnia</b>	<b>People without Insomnia</b>
Asthma	11.6	8.7
Arthritis	39.7	25.4
High Blood Pressure	31.1	24.5
Heart Disease	11.2	8.1
Diabetes	11.0	8.5
Bowel Disorder	14.6	6.6
Migraine	18.9	8.8
Back Problems	38.5	22.0

#### 4.2.2 *Variable Selection*

Three major types of variable selection procedures are widely used in logistic regression: standard, sequential (hierarchical), and statistical (stepwise) regressions. Of the three analytic strategies, stepwise regression produces the best prediction model (Tabachnick & Fidell, 1989). It is a model-building procedure that is useful in predicting the response by choosing the best combination of predictors (Tabachnick & Fidell, 2007). It can be of three forms: forward selection, backward deletion, and stepwise regression. The idea of a stepwise procedure is to include or exclude variables based on statistical significance criteria, which is evaluated via the likelihood chi-square test. “Thus, at any step in the procedure the most important variable, in statistical terms, is the one that produces the greatest change in the log-likelihood relative to a model not containing the variable (i.e., the one that would result in the largest likelihood ratio statistic,  $G$ )” (Hosmer & Lemeshow, 2000).

In this study, stepwise logistic regression was utilized to model insomnia as an outcome. This method was used to determine the most important variables to be included in the final model and to make comparisons between three models that were derived after detecting multivariate outliers (discussed in Section 4.1.4). The stepwise procedure with the significance level of .15 for entry into the model and significance level of .20 to stay in the model was employed to choose a final model, as recommended by Hosmer and Lemeshow (2000). Two statistics, AIC (Akaike Information Criterion) and SC (Schwarz (Bayesian Information) Criterion), are used to compare competing models that are not necessarily nested for the same data. These statistics adjust the  $-2 \log L$  statistic for the number of terms in the model and the number of observations used. Lower values of AIC and SC indicate better fit. These statistics were appropriate to use in this study because the models being

compared had different numbers of observations. Three stepwise logistic procedures were run for Model 1 (a data set with all cases), Model 2 (a data set without the multivariate outliers detected by the method of comparing the leverages of cases with their average leverages), and Model 3 (a data set without the multivariate outliers detected by using the Mahalanobis distance). The procedures of obtaining the three models were demonstrated in Section 4.1.4. Initially, each model had thirteen explanatory variables: AGE, GENDER, MARITAL, PHYSACT, SMOKE, ASTHMA, ARTHRIT, BLOODPR, HEART, DIABETES, BOWEL, MIGRAINE, and BACKPROB. Fit indices and chosen variables for the obtained logistic regression models are presented in Table 9. Based on AIC and SC criteria, Model 2 appears to be the best model.

Table 9. Comparison results for the logistic regression models

<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
<i>Fit Indices</i>	<i>Fit Indices</i>	<i>Fit Indices</i>
DF=15	DF=17	DF=17
AIC=4590.076	AIC=4186.984	AIC=4382.476
SC=4694.213	SC=4302.616	SC=4498.916
<i>Variables</i>	<i>Variables</i>	<i>Variables</i>
AGE	AGE	AGE
GENDER	GENDER	GENDER
SMOKE	SMOKE	SMOKE
ARTHRIT	ARTHRIT	ARTHRIT
	ASTHMA	ASTHMA
BACKPROB	BACKPROB	BACKPROB
BLOODPR	BLOODPR	BLOODPR
BOWEL	BOWEL	BOWEL
	DIABETES	DIABETES
HEART	HEART	HEART
MIGRAINE	MIGRAINE	MIGRAINE

### 4.2.3 Final Multiple Logistic Regression Model

The multiple logistic regression model was fitted against the binary outcome variable INSOMNIA. Eleven explanatory variables were selected in the previous section (Model 2): AGE, GENDER, SMOKE, ARTHRIT, ASTHMA, BACKPROB, BLOODPR, BOWEL, DIABETES, HEART, and MIGRAINE. Maximum likelihood parameter estimates, odds ratios, and goodness-of-fit statistics were calculated from the final fitted logistic regression model.

The final model has the following form:

$$\begin{aligned} \log\left(\frac{\pi}{1-\pi}\right) = & -2.7046 + 0.3300 \text{ GENDER} + 0.3382 \text{ AGE 2} + 0.3971 \text{ AGE 3} \\ & + 0.6104 \text{ AGE 4} + 0.4190 \text{ AGE 5} + 0.0670 \text{ AGE 6} \\ & + 0.00614 \text{ AGE 7} + 0.4269 \text{ SMOKE 2} + 0.1316 \text{ SMOKE 3} \\ & + 0.5272 \text{ ASTHMA} + 0.4135 \text{ ARTHRIT} + 0.2127 \text{ BLOODPR} \\ & + 0.5745 \text{ HEART} + 0.4439 \text{ DIABETES} + 1.3182 \text{ BOWEL} \\ & + 0.9223 \text{ MIGRAINE} + 0.6161 \text{ BACKPROB} \end{aligned} \quad (4.3)$$

The results of the logistic regression are presented in Table 10. The 95% confidence interval for odds ratio for age (groups 6 and 7) and smoke (group 3) contains '1', which indicates that these variables are non-significant. Age group 2 is approaching significance. The odds ratio is interpretable as effect size; the closer it is to '1', the smaller the effect. Bowel disorder (OR = 3.737) followed by the migraine (OR=2.515) has the greatest odds ratio demonstrating the strongest association with insomnia.

“Once a logistic model is formulated, its adequacy is evaluated by a variety of statistical tests and indexes. These include: (a) tests of individual parameter estimates in the



model, (b) tests of the overall model, (c) validation of predicted probabilities, and (d) goodness-of-fit statistic” (Peng, So, Stage, & St. John, 2002), as provided below.

(a) The likelihood ratio test, the Wald statistics, and the Score test assess individual parameter estimates. The Wald test is the simplest. It evaluates each of the coefficients in the model and takes the form of the squared parameter estimate divided by its squared standard error

$$W_j = \frac{B_j^2}{SE_{B_j}^2} . \quad (4.4)$$

The estimated model parameters, standard errors, Wald chi-square, and odds ratios with their 95% intervals are presented in Table 10.

All of the explanatory variables, except AGE 6 (65-74 years old), AGE 7 (over 75 years old), and SMOKE 3 (former daily/former occasional), significantly predict insomnia. The  $p$ -value for AGE 2 (25-34 years old) was 0.0517, which is close to being a significant predictor.

The likelihood ratio test is more powerful than the Wald test and it compares nested models by computing the difference in their  $-2 * \log\text{-likelihood}$  which is distributed as  $\chi^2$  (Tabachnick & Fidell, 2007).

$$\chi^2 = [-2 * \log\text{-likelihood for smaller model} - \\ -2 * \log\text{-likelihood for bigger model}] . \quad (4.5)$$

A series of comparisons of nested models were performed to investigate if removing explanatory variables from the model one by one or in sets increased log-likelihood. They are presented in Section 4.2.4.

Table 10. Logistic Regression Analysis of Insomnia

Variables	Estimate	Standard Error	Wald Chi-Square	Odds Ratio (OR)	95% Confidence Interval for Odds Ratio	
					Lower	Upper
Gender	0.3300	0.0832	15.7239***	1.391	1.182	1.637
Age 2 vs. 1	0.3382	0.1738	3.7859	1.402	0.998	1.972
Age 3 vs. 1	0.3971	0.1639	5.8686*	1.487	1.079	2.051
Age 4 vs. 1	0.6104	0.1609	14.3910***	1.841	1.343	2.524
Age 5 vs. 1	0.4190	0.1655	6.4066**	1.520	1.099	2.103
Age 6 vs. 1	0.670	0.1867	0.1289	1.069	0.742	1.542
Age 7 vs. 1	0.00614	0.1950	0.0010	1.006	0.687	1.475
Cigarette Smoking 2 vs. 1	0.4269	0.1091	15.3148***	1.533	1.238	1.898
Cigarette Smoking 3 vs. 1	0.1316	0.0985	1.7824	1.141	0.940	1.384
Asthma	0.5272	0.1400	14.1769***	1.694	1.288	2.229
Arthritis	0.4135	0.0946	19.0904***	1.512	1.256	1.820
High Blood Pressure	0.2127	0.1013	4.4117*	1.237	1.014	1.509
Heart Disease	0.5745	0.1528	14.1446***	1.776	1.317	2.396
Diabetes	0.4439	0.1483	8.9562***	1.559	1.166	2.085
Bowel Disorder	1.3182	0.1464	81.1015***	3.737	2.805	4.978
Migraine	0.9223	0.1181	61.0233***	2.515	1.996	3.170
Back Problems	0.6161	0.0866	50.5640***	1.852	1.562	2.194
Intercept	-2.7046	0.1437	354.3078***			

* significant at the .05 level

**significant at the .01 level

***significant at the .001 level

The Score test statistic tests the hypothesis that the slope parameters of the explanatory variables are jointly equal to 0 (Stokes, Davis, & Koch, 2000). The small  $p$ -value ( $< .0001$ ) of the Score test in the output rejects the null hypothesis that all slope parameters are equal to zero.

(b) The likelihood ratio test was used to compare the constant-only model with the full model (the constant plus all predictors) to evaluate an overall improvement of the model.

The  $\chi^2(df = 17) = 454.1601$  gives  $p < .0001$ , indicating that the set of explanatory variables reliably predicts insomnia.

(c) Validation of predicted probabilities is evaluated with percentages of correct classifications, Somers'  $D$  statistic, sensitivity, or specificity (Peng et al., 2002). Somers'  $D$  statistic indicates about 16% ( $0.403^2 = 0.162$ ) of shared variance between insomnia and the set of predictors (Tabachnick & Fidell, 2007). Classification was unimpressive with 13.9% sensitivity and high with 98% specificity. *Sensitivity* is the true proportion of positive results that a test elicits when performed on subjects known to have a disease. *Specificity* "is the true proportion of negative results that a test elicits when performed on subjects known to be disease free" (Stokes et al., 2002). The results of the success of prediction are presented in Table 11. At a probability level of .5, the correct classification rate was 80.8%, indicating an overall good success rate. These values show that the predictive ability of the model is adequate.

The area under the receiver operating characteristic (ROC) is a useful measure of the accuracy of predictions. It takes values from 0.5 (indicating random prediction) to 1.0 (indicating perfect prediction) (Tabachnick & Fidell, 2007). The area under the ROC curve,  $c$ , was equal to 0.701, which implies moderately high prediction (see Figure 1).

Table 11. Classification table

Prob. Level	Correct		Incorrect		Correct	Percentages	
	Event	Non- Event	Event	Non- Event		Sensitivity	Specificity
0.06	930	0	3639	0	20.4	100.0	0.0
0.08	917	261	3378	13	25.8	98.6	7.2
0.10	866	715	2924	64	34.6	93.1	19.6
0.12	811	1069	2570	119	41.1	87.2	29.4
0.14	719	1704	1935	211	53.0	77.3	46.8
0.16	666	1991	1648	264	58.2	71.6	54.7
0.18	587	2362	1277	343	64.5	63.1	64.9
0.20	559	2531	1108	371	67.6	60.1	69.6
0.22	487	2725	914	443	70.3	52.4	74.9
0.24	440	2870	769	490	72.4	47.3	78.9
0.26	403	3005	634	527	74.6	43.3	82.6
0.28	365	3093	546	565	75.7	39.2	85.0
0.30	335	3177	462	595	76.9	36.0	87.3
0.32	310	3234	405	620	77.6	33.3	88.9
0.34	282	3292	347	648	78.2	30.3	90.5
0.36	258	3356	283	672	79.1	27.7	92.2
0.38	234	3410	229	696	79.8	25.2	93.7
0.40	213	3453	186	717	80.2	22.9	94.9
0.42	192	3484	155	738	80.5	20.6	95.7
0.44	178	3513	126	752	80.8	19.1	96.5
0.46	168	3530	109	762	80.9	18.1	97.0
0.48	143	3546	93	787	80.7	15.4	97.4
0.50	129	3565	74	801	80.8	13.9	98.0
0.52	120	3581	58	810	81.0	12.9	98.4
0.54	103	3591	48	827	80.8	11.1	98.7
0.56	96	3603	36	834	81.0	10.3	99.0
0.58	92	3606	33	838	80.9	9.9	99.1
0.60	81	3614	25	849	80.9	8.7	99.3
0.62	75	3621	18	855	80.9	8.1	99.5
0.64	71	3623	16	859	80.8	7.6	99.6
0.66	61	3626	13	869	80.7	6.6	99.6
0.68	51	3628	11	879	80.5	5.5	99.7
0.70	46	3632	7	884	80.5	4.9	99.8
0.72	41	3634	5	889	80.4	4.4	99.9
0.74	36	3637	2	894	80.4	3.9	99.9
0.76	32	3637	2	898	80.3	3.4	99.9
0.78	21	3638	1	909	80.1	2.3	100.0
0.80	16	3639	0	914	80.0	1.7	100.0

Table 11 (Continued)

0.82	10	3639	0	920	79.9	1.1	100.0
0.84	7	3639	0	923	79.8	0.8	100.0
0.86	5	3639	0	925	79.8	0.5	100.0
0.88	3	3639	0	927	79.7	0.3	100.0
0.90	1	3639	0	929	79.7	0.1	100.0
0.92	1	3639	0	929	79.7	0.1	100.0
0.94	0	3639	0	930	79.6	0.0	100.0

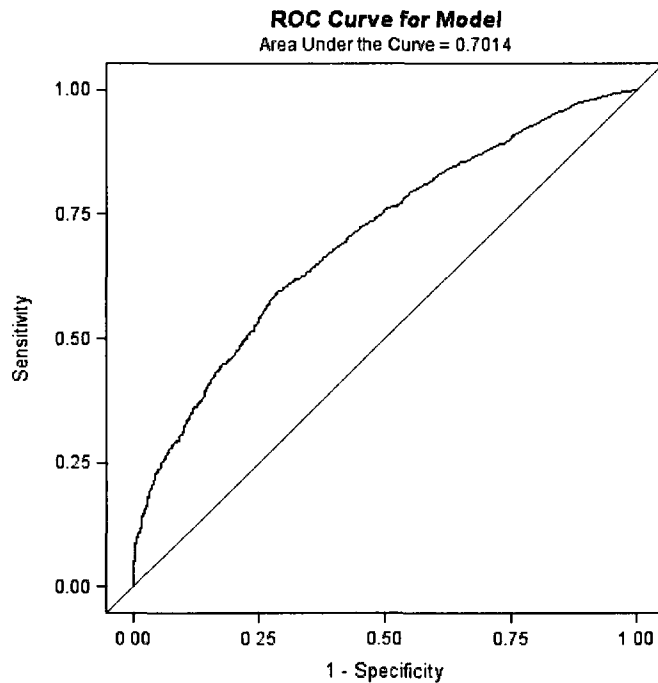


Figure 1. ROC curve for the final model

(d) Once the model is fitted, it is necessary to know how well the model fits the data by assessing the differences between the model-predicted values and the corresponding observed values (Stokes et al., 2000). The Hosmer and Lemeshow Goodness-of-Fit test serves this purpose. In this study, it gives the  $\chi^2$  ( $df=8$ ) = 9.4594,  $p = .3050$ , indicating support of adequacy of the model.

$R^2$  is a measure of model fit and strength of relationship between a set of independent variables and the dependent variable that is widely used in ordinary least squares (OLS). However, in logistic regression  $R^2$  cannot be interpreted as the percent of explained variance in the response (Peng et al., 2002). Therefore, “ $R^2$  type measures can be used to measure relative goodness of fit but may be misleading if used to measure absolute goodness of fit. Models with low values of  $R^2$  can fit great. Models with high values of  $R^2$  can exhibit lack of fit” (Christensen, 1997). There are a number of the  $R^2$  analogues offered in logistic regression. Preference is given to McFadden’s  $R^2$  analogue (Peng et al., 2002), given by:

$$\text{McFadden's } p^2 = 1 - \frac{LL(B)}{LL(O)} , \quad (4.6)$$

where  $LL(B)$  is the log-likelihood for the bigger model (intercept and covariates),  $LL(O)$  is the log-likelihood for the intercept only model (Tabachnick & Fidell, 2007). Here,  $LL(B) = 4,163.076$ ,  $LL(O) = 4,617.236$ , and McFadden’s  $p^2 = 1 - \frac{4,163.076}{4,617.236} = .098$ .

Hosmer and Lemeshow (1989) have reservations about McFadden’s  $p^2$  not being an appropriate goodness-of-fit measure for logistic regression. On the other hand, using Somers’  $D$  statistic equal to 0.403 (see above) as an indicator of model fit shows that there exists a moderately strong and positive association between insomnia and the set of predictors (Healey & Prus, 2009).

#### **4.2.4 Comparing Models**

A series of logistic regressions were run to compare different models listed in Table 12. The comparisons of the models were based on Akaike Information Criterion (AIC) and change in log-likelihood. Logistic regression models, which were run to see whether

removing predictors from the model would change the prediction, are the following: the full model with all demographic, lifestyle and health variables together, the full model with AGE, GENDER and SMOKE removed at each step one by one, then with GENDER and SMOKE removed simultaneously, and finally with demographics and lifestyle variables removed as a set. A lower AIC point corresponds to a better fit. Since the final model with all eleven predictors in it has the lowest AIC = 4,199.076, it verifies that this model is the best.

The  $\chi^2$  statistic was calculated as the difference between the  $[(-2) * \log\text{-likelihood}]$  for the full and reduced models. Degrees of freedom were calculated as the difference between the degrees of freedom for the full and reduced models. All values of  $\chi^2$  were significant, indicating that AGE, GENDER, and SMOKE, one by one or in sets, significantly enhance prediction of insomnia.

Table 12. Comparing models

Model	AIC	-2LogL	$\chi^2$
Full Model: AGE + GENDER + SMOKE + ALL HEALTH FACTORS	4,199.076	4,163.076	
Full Model with AGE removed	4,216.306	4,192.306	29.230***
Full Model with GENDER removed	4,213.017	4,179.017	15.941***
Full Model with SMOKE removed	4,213.703	4,181.703	18.627***
Full Model with AGE and GENDER removed	4,230.416	4,208.416	45.340***
Full Model with AGE, GENDER and SMOKE removed	4,253.175	4,235.175	72.099***
<i>* significant at the .05 level    **significant at the .01 level    ***significant at the .001 level</i>			

Table 13. Odds Ratios for Univariate and Multivariate Logistic Regression Analyses

Effect	Univariate			Multivariate		
	OR	95% CI		OR	95% CI	
Gender	1.553	1.337	1.804	1.391	1.182	1.637
Age 2 vs. 1	1.690	1.227	2.327	1.402	0.998	1.972
Age 3 vs. 1	2.025	1.503	2.727	1.487	1.079	2.051
Age 4 vs. 1	2.656	2.002	3.524	1.841	1.343	2.524
Age 5 vs. 1	2.500	1.886	3.314	1.520	1.099	2.103
Age 6 vs. 1	1.819	1.334	2.481	1.069	0.742	1.542
Age 7 vs. 1	1.704	1.241	2.340	1.006	0.687	1.475
Cigarette Smoking 2 vs. 1	1.918	1.581	2.327	1.533	1.238	1.898
Cigarette Smoking 3 vs. 1	1.412	1.186	1.682	1.141	0.940	1.384
Asthma	1.598	1.245	2.050	1.694	1.288	2.229
Arthritis	2.107	1.811	2.452	1.512	1.256	1.820
High Blood Pressure	1.531	1.306	1.794	1.237	1.014	1.509
Heart Disease	1.969	1.518	2.553	1.776	1.317	2.396
Diabetes	1.635	1.266	2.112	1.559	1.166	2.085
Bowel Disorder	4.858	3.730	6.326	3.737	2.805	4.978
Migraine	2.878	2.332	3.550	2.515	1.996	3.170
Back Problems	2.498	2.141	2.915	1.852	1.562	2.194

Table 13 presents the odds ratios with their confidence intervals for each of the individual predictors. Univariate odds ratios were produced in univariate logistic regression analyses. Then, odds ratios for the individual predictors in the multivariate model were



obtained when all demographic, lifestyle, and health variables were added to the model simultaneously.

The results of univariate analyses indicate that the odds ratios for the variable for age groups 2, 6, and 7 were significant. However, inclusion of all the variables in the model considerably reduced the strong bivariate associations between them and insomnia. On the other hand, the multivariate odds ratio for the variable for asthma increased in comparison with the odds ratio for bivariate association between asthma and insomnia. Bowel disorder followed by migraine demonstrated the strongest associations with insomnia.

### **4.3 Bootstrap**

The bootstrap method was used to draw 10,000 repeated samples of size 4,616 with replacement from the original sample. The 10,000 bootstrap estimates of the odds ratios with their confidence intervals were generated for eleven variables in the final logistic regression model. Also, the bootstrap estimates of the standard errors of the odds ratios were obtained by computing the average of the sample standard deviation of 10,000 values. Odds ratios, their confidence intervals, and standard deviations are presented in Table 14. As can be seen from this table, there were small differences between the odds ratios and their CIs obtained from the single sample and from the 10,000 bootstrap replicates. However, the bootstrap results were consistently, but slightly, larger.

Table 14. Mean values of Odds Ratios, their confidence intervals and standard deviations bootstrap for 10,000 samples

Effect	Mean OR for 10,000				Multivariate		
	replicates		St. Dev.				
	OR	95% CI			OR	95% CI	
Gender	1.401	1.189 1.651	0.118		1.391	1.182 1.637	
Age 2 vs. 1	1.428	1.013 2.013	0.312		1.402	0.998 1.972	
Age 3 vs. 1	1.510	1.092 2.088	0.253		1.487	1.079 2.051	
Age 4 vs. 1	1.876	1.364 2.579	0.312		1.841	1.343 2.524	
Age 5 vs. 1	1.546	1.114 2.145	0.265		1.520	1.099 2.103	
Age 6 vs. 1	1.088	0.752 1.573	0.209		1.069	0.742 1.542	
Age 7 vs. 1	1.026	0.698 1.507	0.206		1.006	0.687 1.475	
Cigarette smoking 2 vs. 1	1.547	1.248 1.919	0.170		1.533	1.238 1.898	
Cigarette smoking 3 vs. 1	1.148	0.946 1.394	0.116		1.141	0.940 1.384	
Asthma	1.713	1.300 2.257	0.237		1.694	1.288 2.229	
Arthritis	1.523	1.264 1.836	0.150		1.512	1.256 1.820	
High Blood Pressure	1.243	1.019 1.518	0.128		1.237	1.014 1.509	
Heart Disease	1.798	1.330 2.430	0.279		1.776	1.317 2.396	
Diabetes	1.577	1.178 2.113	0.236		1.559	1.166 2.085	
Bowel Disorder	3.798	2.844 5.073	0.542		3.737	2.805 4.978	
Migraine	2.538	2.011 3.203	0.303		2.515	1.996 3.170	
Back Problems	1.869	1.575 2.217	0.165		1.852	1.562 2.194	

Table 15. Standard errors for logistic regression parameter estimates and for bootstrapped parameter estimates

<b>Variables</b>	<b>Logistic Regression Parameter Estimates</b>	<b>Bootstrap Parameter Estimates</b>	<b>Logistic Regression Standard Error</b>	<b>Bootstrap Standard Error</b>
Gender	0.3300	0.3338	0.0832	0.0843
Age 2 vs. 1	0.3382	0.3404	0.1738	0.1774
Age 3 vs. 1	0.3971	0.3983	0.1639	0.1658
Age 4 vs. 1	0.6104	0.6154	0.1609	0.1639
Age 5 vs. 1	0.4190	0.4213	0.1655	0.1689
Age 6 vs. 1	0.0670	0.0660	0.1867	0.1900
Age 7 vs. 1	0.0061	0.0056	0.1950	0.1983
Cigarette Smoking 2 vs. 1	0.4269	0.4307	0.1091	0.1093
Cigarette Smoking 3 vs. 1	0.1316	0.1332	0.0985	0.1002
Asthma	0.5272	0.5288	0.1400	0.1386
Arthritis	0.4135	0.4162	0.0946	0.0983
High Blood Pressure	0.2127	0.2126	0.1013	0.1027
Heart Disease	0.5745	0.5746	0.1528	0.1549
Diabetes	0.4439	0.4446	0.1483	0.1500
Bowel Disorder	1.3182	1.3245	0.1464	0.1416
Migraine	0.9223	0.9242	0.1181	0.1193
Back Problems	0.6161	0.6213	0.0866	0.0882
Intercept	-2.7046	-2.7199	0.1437	0.1427

Table 15 presents the standard errors of parameter estimates for the logistic regression model as well as the bootstrap standard errors. Based on these values it can be observed that

the parameter estimates are similar for both classical and bootstrap logistic regression models, however, slightly higher for the latter, except for the variables for age groups 6 and 7 and high blood pressure. In addition, it can be noted that standard errors are slightly higher for the bootstrap parameter estimates, except for the variables for asthma, bowel disorder, and the intercept. It is noted that the bootstrap estimates have somewhat higher standard errors in comparison with the classical logistic regression estimates.

The PROC SURVEYSELECT procedure in SAS 9.2 was used to obtain 10,000 replicates when sampling with replacement by the method of URS (Unrestricted Random Sample) with the random seed of 711765001. The seed used is important information if there is a need to reproduce the results (Cassell, 2007).

Histograms for 10,000 bootstrap samples for odds ratio for each of the eleven variables were produced. From the histograms in Figures 2 and 3 below and Figures 4-18 in Appendix C, it can be observed that the distributions appear approximately normal, however, slightly skewed. The results of the Kolmogorov-Smirnov test for normality indicated that these distributions were not normal.

The sampling distributions of odds ratios were estimated with the empirical distribution function. Fitting an empirical (parametric) distribution that represents data in a better way is a key in statistical analyses (Curtis, 2007). A gamma distribution was empirically chosen to estimate the distributions of the odds ratios for the predictors. The parameters of gamma distributions are  $\alpha$  and  $\beta$ . The estimates of parameters along with the values for the mean and variance of the distributions are presented in Table 16.

Table 16. Parameters of gamma distribution, its mean and variance

Variable	$A$	$\beta$	Mean $\mu=\alpha \beta$	Variance $\sigma^2=\alpha\beta^2$
Gender	1.567	0.894	1.401	0.014
Age 2 vs. 1	21.000	0.068	1.428	0.066
Age 3 vs. 1	35.952	0.042	1.510	0.064
Age 4 vs. 1	36.077	0.052	1.876	0.097
Age 5 vs. 1	34.356	0.045	1.546	0.070
Age 6 vs. 1	27.200	0.040	1.088	0.044
Age 7 vs. 1	25.024	0.041	1.026	0.042
Cigarette smoking 2 vs. 1	81.421	0.019	1.547	0.029
Cigarette smoking 3 vs. 1	95.667	0.012	1.148	0.014
Asthma	51.909	0.033	1.713	0.056
Arthritis	101.533	0.015	1.523	0.023
High Blood Pressure	95.615	0.013	1.243	0.016
Heart Disease	41.814	0.043	1.798	0.078
Diabetes	45.057	0.035	1.577	0.056
Bowel Disorder	49.325	0.077	3.798	0.294
Migraine	70.500	0.036	2.538	0.092
Back Problems	124.600	0.015	1.869	0.027

The choice for a gamma distribution was made based on empirical knowledge. However, there is a procedure in SAS that allows determining an appropriate model for a data distribution by considering curves from several distribution families. It was decided to

run it to see whether or not the results using computer power will differ from those obtained by using empirical knowledge.

PROC UNIVARIATE procedure in SAS enables to fit a wide range of parametric distributions including normal, lognormal, gamma, and Weibull, and visualize them through graphs. In addition, PROC UNIVARIATE provides a series of goodness-of-fit tests and  $p$ -values based on the empirical distribution function (EDF). The EDF tests are Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Mises statistics based on the difference between the empirical and cumulative distribution functions. The goodness-of-fit test statistics are computed to test the null hypothesis that the values of the variables in the analysis are a random sample from the specified theoretical distribution (Curtis, 2007). The probability values for the EDF tests for normal and Weibull models were all less than .10, indicating that the data do not support either of these models.

The choice for gamma distribution was confirmed for the following variables: ASTHMA, BACKPROB (log-normal is also supported for this variable), BLOODPR, DIABETES, and MIGRAINE. At the  $\alpha = .10$  significance level, all tests supported the conclusion that the two-parameter gamma distribution with the shape parameter  $\alpha$  and the scale parameter  $\beta$  provided a good model for the distribution of the odds ratios for these variables (Table 17).

As for the variables AGE (except AGE 4), ARTHRIT, BOWEL, GENDER, and SMOKE 2, at the  $\alpha = .10$  significance level, all tests supported the conclusion that the two-parameter log-normal distribution with the scale parameter  $\zeta$  and the shape parameter  $\sigma$  provide a good model for the distributions of the odds ratios for these variables (Table 18).

Neither of the parametric distributions based on  $\alpha = .10$  significance level EDF tests provided a good model for AGE 4, SMOKE 3, and HEART variables.

Table 17. Parameters of gamma distribution, its mean and variance

<b>Variable</b>	<b>$A$</b>	<b>$\beta$</b>	<b>Mean $\mu=\alpha \beta$</b>	<b>Variance $\sigma^2=\alpha\beta^2$</b>
Asthma	52.526	0.033	1.713	0.056
High Blood Pressure	95.278	0.013	1.243	0.016
Diabetes	44.819	0.035	1.577	0.056
Migraine	70.570	0.036	2.538	0.092
Back Problems	128.692	0.015	1.869	0.027

Table 18. Parameters of lognormal distribution, its mean and variance

<b>Variable</b>	<b>$\sigma$</b>	<b>$\xi$</b>	<b>Mean</b>	<b>Variance</b>
Gender	0.084	0.334	1.401	0.014
Age 2 vs. 1	0.177	0.340	1.428	0.065
Age 3 vs. 1	0.166	0.398	1.510	0.064
Age 5 vs. 1	0.169	0.421	1.546	0.069
Age 6 vs. 1	0.190	0.066	1.088	0.043
Age 7 vs. 1	0.198	0.006	1.026	0.042
Cigarette smoking 2 vs. 1	0.109	0.431	1.547	0.029
Arthritis	0.098	0.416	1.523	0.023
Bowel Disorder	0.142	1.324	3.798	0.292

The graphs for both log-normal and gamma distributions were produced and are exhibited in Figures 2 and 3 below and in Figures 4-18 presented in Appendix C. Visually,

there seems to be a slight difference between them. The solid line represents the log-normal distribution, whereas the dotted line represents the gamma distribution.

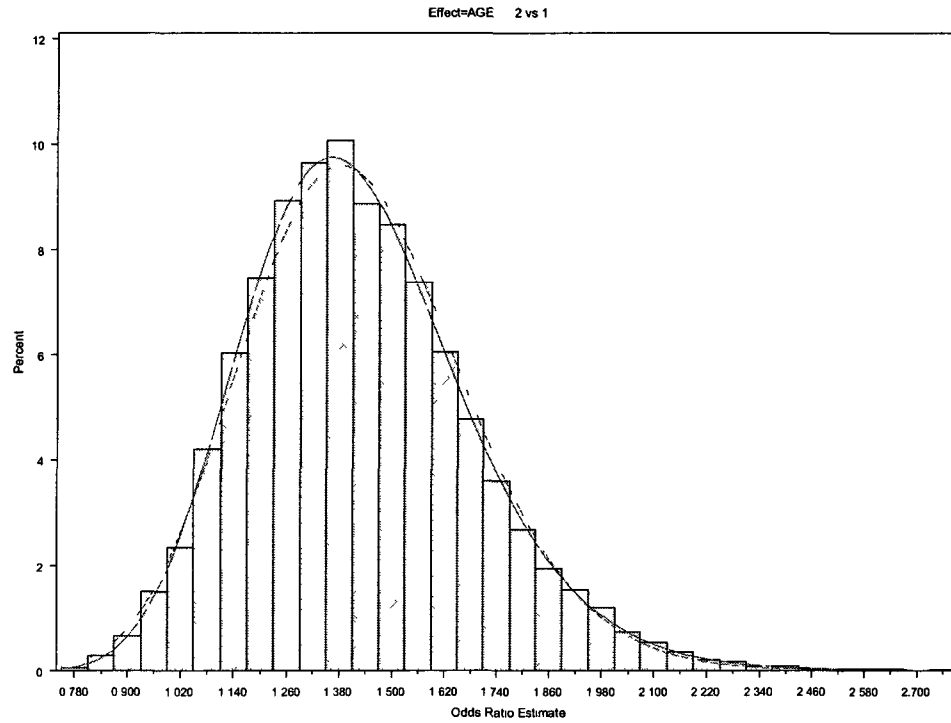


Figure 2. Histogram for AGE 2 vs. 1



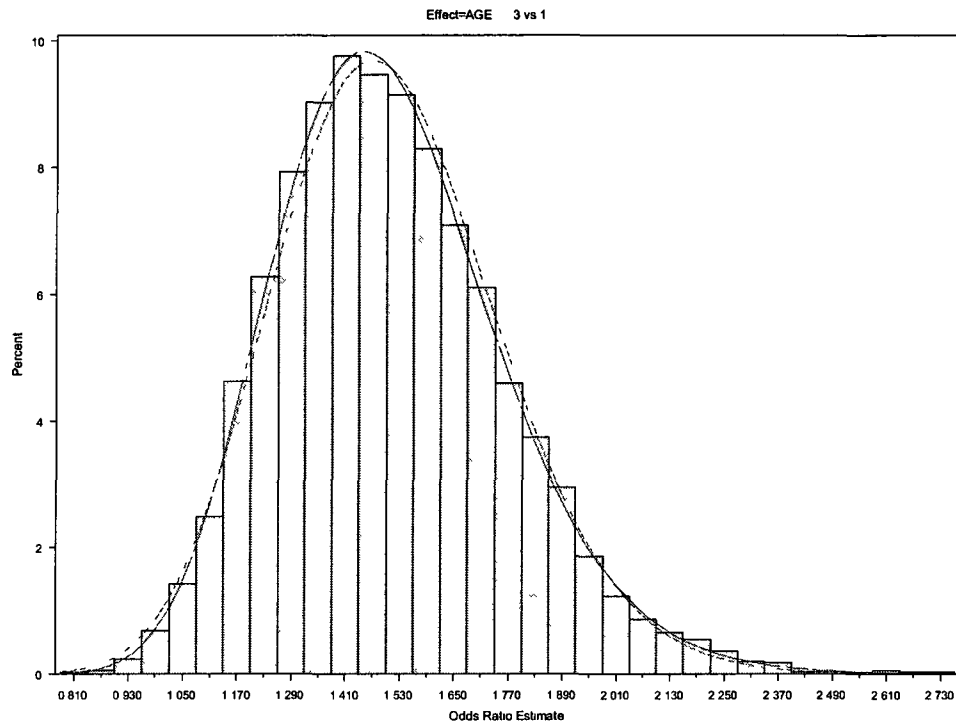


Figure 3. Histogram for AGE 3 vs. 1

## **CHAPTER 5 – DISCUSSION AND CONCLUSIONS**

### **5.1 Discussion**

At the time of writing, this study was the only one to use the Canadian Community Health Survey conducted in the years 2007-2008 to examine the association between potential risk factors and insomnia in the province of Nova Scotia. Emphasis was placed on assessing the relationship between insomnia and health conditions, such as asthma, arthritis, migraine, heart disease, high blood pressure, diabetes, bowel disorder, and back problems, excluding fibromyalgia and arthritis. Series of logistic regression analyses were performed to make this assessment. The bootstrap method was then applied to obtain odds ratios with their confidence intervals for the predictors of insomnia.

The results of this study reveal that age 65 and over did not show a significant association with insomnia after controlling for demographic and lifestyle characteristics and health variables. Sutton et al. (2001) as well did not find a significant association between insomnia and age, including all age groups, which led the authors to conclude that age is not a significant risk factor for insomnia. Dregan and Armstrong (2009) confirmed these results of such association in adults aged 50 and older and generalized from facts an explanation for this finding as follows: the age-related changes in sleep disruption are caused by medical and psychiatric diseases and other human risk factors (e.g., financial strain). On the contrary, Katz and McHorney (1998) demonstrated that advanced age (i.e., > 65) was highly associated with insomnia both in bivariate analysis and after adjusting for chronic conditions. The results of other investigators of associations between insomnia and female gender are extended by the

findings of this study, confirming that female population relates to higher risk for insomnia (Katz & McHorney, 1998; Roberts et al., 2008; Sutton et al., 2001; Su et al., 2004). Gellis et al. (2005) also found that risk for insomnia was significantly greater with increasing age and among women.

Katz and McHorney (1998) did not observe a significant association between smoking status and insomnia. In this study, however, an association was found between insomnia and the current level of smoking (those who smoke daily and occasionally (former daily smokers) and those who have always been occasional smokers). Patten et al. (2000) found cigarette smoking to be a significant predictor for the increased risk of developing of frequent sleep problems. Marital status was not associated with insomnia in the current study. Sutton et al. (2001) on the other hand, found significant association between insomnia and being widowed or single. This result was also reported in an earlier study by Katz and McHorney (1998). Su et al. (2004) found high association between insomnia and single status in men, while this association was not significant for women.

Physical activity was not found to be a risk factor for insomnia in this study. This result contradicts the findings of Morgan (2003), who demonstrated that a lower physical activity level was a significant risk factor for insomnia among people of age 65 and older. Education did not have a significant association with insomnia in this study. Gellis et al. (2005), on the other hand, found education to be a risk factor of insomnia.

An association between insomnia and several medical problems has also been demonstrated (Katz & McHorney, 1998; Power et al., 2005; Sutton et al., 2001; Taylor et al., 2007). Some of these results are in close agreement with each other and are consistent with the findings of this study. Bowel disorder showed the highest association with insomnia in this study, which is in contrast to the findings of Katz and McHorney (1998). They note that

insignificance of their results may be explained by the incompleteness of the study questionnaires.

The second highest association between migraine and insomnia symptoms was observed in the present study. This result is also confirmed in the studies conducted by Power et al. (2005) and Sutton et al. (2001). The result of the present study reveals a strong association between insomnia and back problems. Back problems were also reported as risk factors for mild and severe insomnia by Katz and McHorney (1998).

People with heart disease, high blood pressure, and diabetes were more likely to have insomnia according to the results presented in this study. These findings are in accordance with the results of Taylor et al. (2007) and Power et al. (2005). On the other hand, diabetes was not significantly associated with insomnia in the study of Sutton et al. (2001).

In addition to the logistic regression analysis carried out in this thesis, the bootstrap method was used to obtain more reliable estimates of odds ratios and their 95% confidence intervals in addition to the standard errors. The bootstrap parameter estimates along with their standard errors were observed to be slightly larger, on average, than the logistic regression parameter estimates and standard errors. It should be noted that the absolute average difference between such values was very small.

All in all, the power of any statistical analysis increases with sample size. In this study, the data set is large, and the number of explanatory variables in the final logistic regression model is equal to 11, so there is no issue with the ratio of cases to explanatory variables. Tabachnick and Fidell (2007) recommend having at least 5 times more cases than explanatory variables as a bare minimum requirement. When stepwise regression is used, this number is higher (i.e., the ratio of cases to explanatory variables should be 40 to 1).

When there is a large sample size (5,018 in this study), very small differences will be detected as significant. Therefore, it is reasonable to use a decreased significance level for hypothesis testing. A significance level of .01 and .001 was used for the  $\chi^2$  tests in this thesis, and the conventional 5% significance level (which is the default option in SAS) was used for logistic regression analyses.

## **5.2 Conclusions**

This study intended to assess the relationship between insomnia and health factors, demographics, socioeconomic, and lifestyle characteristics. According to the results, among demographic and lifestyle factors, age younger than 65 years old, female gender, and current level of smoking were found to have significant associations with insomnia. Arthritis, asthma, back problems, high blood pressure, bowel disorder, diabetes, heart disease, and migraine were found to be significant risk factors among health factors for insomnia in the population of Nova Scotia.

## **5.3 Limitations**

There are some limitations in this study. First, the reliability of self-reported data can be doubted. The questions about sleep were based on retrograde details, particularly taking into account how often individuals had trouble going to sleep or staying asleep. Some questions covered a time frame of the past 12 months. In addition, it is not known if the subjects had current insomnia or if their insomnia was in the chronic state, as well as whether insomnia was primary or secondary. Moreover, the CCHS did not contain questions

regarding taking medications to cure medical conditions in patients experiencing them or consuming caffeine. Use of some of the medications can have insomnia as a side effect (Taylor et al., 2007), as well as caffeine intake has been known to influence sleep.

Also, several epidemiological studies on insomnia have examined the relationship between insomnia and stress, depression, anxiety, and other psychological problems. This relationship is firmly established in the literature (Koffel & Watson, 2009; LeBlanc et al., 2007). The abovementioned variables, however, were not included in the analyses conducted in this thesis; though, the inclusion of them might have provided confounding effect in the relationship between insomnia and health, demographic, and lifestyle variables. Finally, the direction of the cause-effect relationship between insomnia and its associated factors was not established because the cross-sectional nature of the CCHS precludes causal conclusions.

## **5.4 Implications and Recommendations for Future Work**

The findings of this study can contribute to future research on sleep problems and associated chronic illnesses that are prevalent in the province of Nova Scotia. Given that sleep was an optional content in the cross-sectional Canadian Community Health Survey as described in Section 3.3, the questions about sleep were brief and based on recalling details of the events happening during the twelve-month period. The accuracy of such self-assessment can be questioned. That is why a longitudinal study that is focused exclusively on insomnia needs to be conducted in the province of Nova Scotia, and the effects of depression, anxiety, and other psychological conditions, together with chronic illnesses should be accounted for. It is worth mentioning that there is a need for an accurate definition of insomnia, so that the quality of sleep of participants in studies can be determined.

Consequently, more precise conclusions can be drawn. This study can now be employed to make comparisons with other studies conducted on insomnia not only in Canada, but all over the world.

Reporting of these results to the health care providers in the province of Nova Scotia will shed light on the problem of insomnia in that region, so that necessary steps are taken to conduct further research. In turn, this will allow researchers to conduct a thorough investigation of sleep problems in the province by carefully examining the symptoms and taking into consideration the duration, chronicity, and the nature (primary or secondary) of sleep problems. It should be highlighted that in order to improve the quality of life of people who suffer from insomnia, it should not be ignored, but recognized as a serious medical condition.

The results of this study might raise the attention of researchers and practitioners and encourage them to take measures in viewing insomnia not only as a symptom of other medical conditions, but as a disease that needs to be addressed. In this context, insomnia has been found to be associated with significant mortality in men. Therefore, it should be identified at an earlier stage so that it can be treated properly. Moreover, insomnia in children is found to be related to mental disease; therefore, this suggests the need for the development of an early intervention programs to identify the symptoms and initiate measures of treatment (Matsuyama & Cortez, 2010).

## Bibliography

- Abraham, B., & Ledolter, J. (2005). *Introduction to regression modeling*. Pacific Grove, CA: Duxbury Press.
- Akerstedt, T. (2000). Consensus statement: Fatigue and accidents in transport operations. *Journal of Sleep Research*, 9 (4), 395.
- Ancoli-Israel, S. (2006). The impact and prevalence of chronic insomnia and other sleep disturbances associated with chronic illness. *The American Journal of Managed Care*, 12 (8), S221-S229.
- Barker, N. A. (2005). *A practical introduction to the bootstrap using the SAS system*. Wallingford, UK: Oxford Pharmaceutical Sciences.
- Beirness, D. J., Simpson, H. M., & Desmond, K. (2005). *The road safety monitor 2004: Drowsy driving*. Ottawa, Ontario: Traffic Injury Research Foundation.
- Capital Health Nova Scotia. (2008). *Why change the way we think about health?* Retrieved June 25, 2010, from <http://www.cdha.nshealth.ca/>
- Cassell, D. L. (2007). Don't be loopy: Re-sampling and simulation the SAS way. *SAS Global Forum 2007 Statistics and Data Analysis*. SAS Inc.
- Christensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). New York, NY: Springer-Verlag.
- Curtis, N. A. (2007). *Are histograms giving you fits? New SAS software for analyzing distributions*. Retrieved 2010, from <http://www.ats.ucla.edu/stat/sas/library/distributionanalysis.pdf>
- Davidson, J. R., MacLean, A. W., Brundage, M. D., & Schulze, K. (2002). Sleep disturbance in cancer patients. *Social Science & Medicine*, 54 (9), 1309-1321.
- Dregan, A., & Armstrong, D. (2009). Age, cohort and period effects in the prevalence of sleep disturbances among older people: The impact of economic downturn. *Social Science & Medicine*, 69 (10), 1432-1438.
- Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- Everitt, B. S., & Hothorn, T. (2006). *A handbook of statistical analyses using R*. London, UK: Chapman & Hall/CRC.



- Friendly, M. (2000). *Visualizing categorical data*. Cary, NC: SAS Institute Inc.
- Gellis, L. A., Lichstein, K. L., Scarinci, I. C., Durrence, H. H., Taylor, D. J., Bush, A. J., et al. (2005). Socioeconomic status and insomnia. *Journal of Abnormal Psychology, 114* (1), 111-118.
- Hajak, G. (2001). Epidemiology of severe insomnia and its consequences in Germany. *European Archives of Psychiatry and Clinical Neuroscience, 251* (2), 49-56.
- Harrell Jr., F. E. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. New York, NY: Springer-Verlag.
- Healey, J. F., & Prus, S. G. (2009). *Statistics: A tool for social research*. Scarborough, ON: Nelson Education Ltd.
- Helpguide. (2010). *Sleep disorders and problems*. Retrieved April 15, 2010, from [http://www.helpguide.org/life/sleep_disorders.htm](http://www.helpguide.org/life/sleep_disorders.htm)
- Highway Safety Roundtable. (2007). *Fatigue impairment: Police issues*. Retrieved May 20, 2010, from [http://www.fatigueimpairment.ca/documents/2007_08_16_fatigueimpairmentpoliceissues.pdf](http://www.fatigueimpairment.ca/documents/2007_08_16_fatigueimpairmentpoliceissues.pdf)
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.
- Howell, D. C. (2009, March 7). Retrieved December 23, 2010, from Treatment of missing data: [http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html](http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html)
- Hurlburt, R. T. (2005). *Comprehending behavioral statistics* (4th ed.). Florence, KY: Wadsworth Publishing.
- Johnson, E. O., Roth, T., Schultz, L., & Breslau, N. (2006). Epidemiology of DSM-IV insomnia in adolescence: Lifetime prevalence, chronicity, and an emergent gender difference. *Pediatrics, 117* (2), e247-e256.
- Katz, D. A., & McHorney, C. A. (1998). Clinical correlates of insomnia in patients with chronic illness. *Archives of Internal Medicine, 158* (10), 1099-1107.
- Katz, D. A., & McHorney, C. A. (2002). The relationship between insomnia and health-related quality of life in patients with chronic illness. *The Journal of Family Practice, 51* (3), 229-236.

- Kim, K., Uchiyama, M., Liu, X., Shibui, K., Ohida, T., Ogihara, R., et al. (2001). Somatic and psychological complaints and their correlates with insomnia in the Japanese general population. *Psychosomatic Medicine*, 63 (3), 441-446.
- Koffel, E., & Watson, D. (2009). The two-factor structure of sleep complaints and its relation to depression and anxiety. *Journal of Abnormal Psychology*, 118 (1), 183-194.
- LeBlanc, M., Beaulieu-Bonneau, S., Merette, C., Savard, J., Ivers, H., & Morin, C. M. (2007). Psychological and health-related quality of life factors associated with insomnia in a population-based sample. *Journal of Psychosomatic Research*, 63 (2), 157-166.
- Leger, D., Guilleminault, C., Bader, G., Levy, E., & Paillard, M. (2002). Medical and socio-professional impact of insomnia. *Sleep*, 25 (6), 621-625.
- Lichstein, K. L., Durrence, H. H., Riedel, B. W., & Bayen, U. J. (2001). Primary versus secondary insomnia in older adults: Subjective sleep and daytime functioning. *Psychology and Aging*, 16 (2), 264-271.
- Linton, S. J. (2004). Does work stress predict insomnia? A prospective study. *British Journal of Sleep Psychology*, 9 (2), 127-136.
- Liu, X., Uchiyama, M., Kim, K., Okawa, M., Shibui, K., Kudo, Y., et al. (2000). Sleep loss and daytime sleepiness in the general adult population of Japan. *Psychiatry Research*, 93 (1), 1-11.
- Mai, E., & Buysse, D. J. (2008). Insomnia: Prevalence, impact, pathogenesis, differential diagnosis, and evaluation. *Sleep Medicine Clinics*, 3 (2), 167-174.
- Matsuyama, K., & Cortez, M. F. (2010, September 1). Retrieved February 5, 2011, from Insomnia triggers men's death, kids' mental decline: <http://www.businessweek.com/news/2010-09-01/insomnia-triggers-men-s-death-kids-mental-decline.html>
- Mitler, M. M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1988). Catastrophes, Sleep, and Public Policy: Consensus Report. *Sleep*, 11 (1), 100-109.
- Moore, D. S., & McCabe, G. P. (2005). Introduction to the practice of statistics. W. H. Freeman.
- Morgan, K. (2003). Daytime activity and risk factors for late-life insomnia. *The Journal of Sleep Research*, 12 (3), 231-238.

- National Sleep Foundation. (2010). *Can't sleep? What to know about insomnia*. Retrieved April 21, 2010, from <http://www.sleepfoundation.org/article/sleep-related-problems/insomnia-and-sleep>
- Ohayon, M. M., & Bader, G. (2010). Prevalence and correlates of insomnia in the Swedish population aged 19-75 years. *Sleep Medicine, 11* (10), 980-986.
- Ohayon, M. M., & Smirne, S. (2002). Prevalence and consequences of insomnia disorders in the general population of Italy. *Sleep Medicine, 3* (2), 115-120.
- Ohayon, M. M., Roberts, R. E., Zulley, J., Smirne, S., & Priest, R. G. (2000). Prevalence and patterns of problematic sleep among older adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 39* (12), 1549-1556.
- Okun, M. L., Kravitz, H. M., Sowers, M. F., Moul, D. E., Buysse, D. J., & Hall, M. (2009). Psychometric evaluation of the insomnia symptom questionnaire: A self-report measure to identify chronic insomnia. *Journal of Clinical Sleep Medicine, 5* (1), 41-51.
- Patlak, M. (2005). *Your guide to healthy sleep*. US Department of Health and Human Services. NIH Publication No. 06-5271.
- Patten, C. A., Choi, W. S., Gillin, J. C., & Pierce, J. P. (2000). Depressive symptoms and cigarette smoking predict development and persistence of sleep problems in US adolescents. *Pediatrics, 102* (2), 1-9.
- Peng, C.-Y. J., So, T.-S. H., Stage, F. K., & St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999. *Research in Higher Education, 43* (3), 259-293.
- Philip, P., & Akerstedt, T. (2006). Transport and industrial safety, how are they affected by sleepiness and sleep restriction? *Sleep Medicine Reviews, 10* (5), 347-356.
- Power, J. D., Perruccio, A. V., & Badley, E. M. (2005). Pain as a mediator of sleep problems in arthritis and other chronic conditions. *Arthritis & Rheumatism (Arthritis Care & Research), 53* (6), 911-919.
- Riedel, B. W., & Lichstein, K. L. (2000). Insomnia and daytime functioning. *Sleep Medicine Reviews, 4* (3), 277-298.
- Riedel, B. W., Durrence, H. H., Lichstein, K. L., Taylor, D. J., & Bush, A. J. (2004). The relation between smoking and sleep: the influence of smoking level, health, and psychological variables. *Behavioral Sleep Medicine, 2* (1), 63-78.

- Roberts, R. E., Roberts, C. R., & Duong, H. T. (2008). Chronic insomnia and its negative consequences for health and functioning of adolescents: A 12-month prospective study. *Journal of Adolescent Health, 42* (3), 294-302.
- Rosekind, M., Gander, P., Gregory, K., Smith, R., Miller, D., Oyung, R., et al. (1996). Managing fatigue in operational settings 1: Physiological considerations and countermeasures. *Behavioral Medicine, 21* (4), 157-165.
- Roth, T. (2007). Insomnia: definition, prevalence, etiology, and consequences. *Journal of Clinical Sleep Medicine, 3* (S5), S7-S10.
- Schwab, A. J. (2002, February 15). *Analyzing patterns of missing data*. Retrieved December 3, 2010, from [http://www.utexas.edu/courses/schwab/sw388r7/Tutorials/Analyzing_Patterns_of_Missing_Data_doc_html/001_Analyzing_Patterns_of_Missing_Data.html](http://www.utexas.edu/courses/schwab/sw388r7/Tutorials/Analyzing_Patterns_of_Missing_Data_doc_html/001_Analyzing_Patterns_of_Missing_Data.html)
- Statistics Canada. (2009). *Canadian Community Health Survey (CCHS), user guide 2007-2008 microdata files*. Health Statistics Division. Ottawa: Statistics Canada.
- Stein, M. D., & Friedmann, P. D. (2005). Disturbed sleep and its relationship to alcohol use. *Journal of Substance Abuse Treatment, 26* (1), 1-13.
- Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system*. Cary, NC: SAS Institute Inc.
- Stoller, M. K. (1997). The socio-economics of insomnia: The materials and the methods. *European Psychiatry, 12* (S1), 41-48.
- Su, T.-P., Huang, S.-R., & Chou, P. (2004). Prevalence and risk factors of insomnia in community-dwelling Chinese elderly: A Taiwanese urban area survey. *Australian and New Zealand Journal of Psychiatry, 38* (9), 706-713.
- Sutton, D. A., Moldofsky, H., & Badley, E. M. (2001). Insomnia and health problems in Canadians. *Insomnia, 24* (6), 665-670.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York, NY: HarperCollins Publishers, Inc.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.
- Taylor, D. J., Lichstein, K. L., & Durrence, H. H. (2003). Insomnia as a health risk factor. *Behavioral Sleep Medicine, 1* (4), 227-247.

- Taylor, D. J., Mallory, L. J., Lichstein, K. L., Durrence, H. H., Riedel, B. W., & Bush, A. J. (2007). Comorbidity of chronic insomnia with medical problems. *Sleep, 30* (2), 213-218.
- Turkoski, B. B. (2006). Managing insomnia. *Pharmacology, 25* (5), 339-345.
- Vinson, D. C., Manning, B. K., Galliher, J. M., Dickinson, L. M., Pace, W. D., & Turner, B. J. (2010). Alcohol and sleep problems in primary care patients: A report from the AAFP national research network. *Annals of Family Medicine, 8* (6), 484-492.
- Zhang, B., & Wing, Y.-K. (2006). Sex differences in insomnia: A meta-analysis. *Sleep, 29* (1), 85-93.

## **Appendix A – List of Variables in CCHS 2007-2008**

In this appendix, we illustrate the variables used in this study that were chosen from the Statistics Canada file.

### **SLP_02 – Trouble sleeping**

- 1-None of the time
- 2- Little/time
- 3-Some of the time
- 4-Most of the time
- 5-All of the time
- 6-Not applicable
- 7-Don't know
- 8-Refusal
- 9-Not stated

### **DHH_SEX – Gender**

- 1-Male
- 2-Female

### **DHHGMS- Marital status**

- 1-Married
- 2-Common-law
- 3-Widdow/Separated/Divorced
- 4-Single/Never married
- 9-Not stated

**DHHGAGE – Age**

1-12 to 14 years

2-15 to 17 years

3-18 to 19 years

4-20 to 24 years

5-25 to 29 years

6-30 to 34 years

7-35 to 39 years

8-40 to 44 years

9-45 to 49 years

10-50 to 54 years

11-55 to 59 years

12-60 to 64 years

13-65 to 69 years

14-70 to 74 years

15-75 to 79 years

16-80 years or more

**ALCDDTM- Type of drinker (12 months)**

1-Regular drinker

2-Occasional drinker

3-No drink last 12 months

4-Not stated

**EDUDH04- Highest level of education - household**

- 1-Less than secondary school graduation
- 2-Secondary school graduation
- 3-Some post-secondary
- 4-Post-secondary graduation
- 9-Not stated

**INCGHH-Total household income from all sources**

- 1-No income or less than \$20,000
- 2-\$20,000 to \$39,999
- 3-\$40,000 to \$59,999
- 4-\$60,000 to \$79,999
- 5-\$80,000 or more
- 9-Not stated

**SMKDSTY-Type of smoker**

- 1-Daily smoker
- 2-Occasional smoker (former daily smoker)
- 3-Always an occasional smoker
- 4-Former daily smoker
- 5-Former occasional smoker
- 6-Never smoked
- 99-Not stated



**PACDPAL-Leisure physical activity index**

1-Active

2-Moderately active

3-Inactive

9-Not stated

**CCC_031- Has asthma**

1-Yes

2-No

3-Don't know

8-Refusal

9-Not stated

**CCC_91A- Has chronic bronchitis**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_91E-Has emphysema (Respondents aged 30 and over)**

1-Yes

2-No

6-Not applicable

7-Don't know

8-Refusal

9-Not stated

**CCC_051-Has arthritis (excluding fibromyalgia)**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_071-Has high blood pressure**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_121-Has heart disease**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_101-Has diabetes**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_141-Has stomach or intestinal ulcers**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_171-Has a bowel disorder/Colitis**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_081-Has migraine headaches**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_131-Has cancer**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

**CCC_061-Has back problems excluding fibromyalgia and arthritis**

1-Yes

2-No

7-Don't know

8-Refusal

9-Not stated

DSM-IV symptom criteria to diagnose insomnia include:

1. Difficulty initiating asleep (DIS)
2. Difficulty maintaining sleep (DMS 1&2, two items)
3. Early morning awakening (EMA)
4. Non –restorative sleep (NRS)

The time frame is 4 weeks.

## Appendix B – Recoding Designations for the Variables in this Study

The following table lists all the variables used in the study, their labels, types, and recoded designations.

Table 19. Recoding designations for the variables

New variable	Label	Variable Type	Coding	Statistics Canada Variable Name
INSOMNIA	Trouble sleeping	Dichotomous	0=absence of sleep problems 1=presence of sleep problems .=missing	Slp_02=1, 2 or 3 Slp_02=4 or 5 Slp_02=7, 8 or 9
GENDER	Sex of respondent	Dichotomous	0=male 1=female	Dhh_sex=1 Dhh_sex=2
AGE	Age of respondent	Ordinal	1=12-24 2=25-34 3=35-44 4=45-54 5=55-64 6=65-74 7=75+	Dhhgage=1, 2, 3, 4 Dhhgage=5, 6 Dhhgage=7, 8 Dhhgage=9, 10 Dhhgage=11, 12 Dhhgage=13,14 Dhhgage=15, 16
MARITAL	Marital status	Ordinal	1=married/common-law 2=widowed/separated/divorced 3=single/never married .=missing	DHHGMS=1, 2 DHHGMS=3 DHHGMS=4 DHHGMS=9
ALCOHOL	Type of drinker	Ordinal	1=no drink last 12 months 2=regular drinker 3=occasional drinker .=missing	ALCDTTM=3 ALCDTTM=1 ALCDTTM=2 ALCDTTM=9

*Table 19 (Continued)*

EDUCAT	Level of education	Ordinal	1=< than secondary 2=secondary grad. 3=post-sec. grad 4=other post secondary .=missing	EDUDR04=1 EDUDR04=2 EDUDR04=3 EDUDR04=4 EDUDR04=9
INCOME	Total income	Ordinal	1= \$80,000 or more 2= no or <\$20,000 3= \$20,000-\$39,999 4= \$40,000-\$59,999 5= \$60,000-\$79,999 .=missing	INCGHH=5 INCGHH=1 INCGHH=2 INCGHH=3 INCGHH=4 INCGHH=6
PHYSACT	Physical activity index	Ordinal	1=active 2=moderately active 3=inactive .=missing	PACDPAI=1 PACDPAI=2 PACDPAI=3 PACDPAI=9
SMOKE	Cigarette smoking	Ordinal	1=never smoked 2=current (daily/ occasional/ always occasional) 3=former (former daily/ former occasional) .=missing	SMKDSTY=6 SMKDSTY=1, 2, 3 SMKDSTY=4, 5 SMKDSTY=99
ASTHMA	Asthma	Dichotomous	0=absence 1=presence	CCC_031=2 CCC_031=1
BRONCH	Chronic bronchitis	Dichotomous	0=absence 1=presence .=missing	CCC_91A=2 CCC_91A=1 CCC_91A=7, 9
EMPHYS	Emphysema	Dichotomous	0=absence 1=presence .=missing	CCC_91E=2 CCC_91E=1 CCC_91E=6, 7
ARTHRIT	Arthritis	Dichotomous	0=absence 1=presence .=missing	CCC_051=2 CCC_051=1 CCC_051=7

---

*Table 19 (Continued)*

BLOODPR	High blood pressure	Dichotomous	0=absence 1=presence .=missing	CCC_071=2 CCC_071=1 CCC_071=7
HEART	Heart disease	Dichotomous	0=absence 1=presence .=missing	CCC_121=2 CCC_121=1 CCC_121=7
DIABETES	Diabetes	Dichotomous	0=absence 1=presence .=missing	CCC_101=2 CCC_101=1 CCC_101=7
ULCERS	Intestinal or stomach ulcers	Dichotomous	0=absence 1=presence .=missing	CCC_141=2 CCC_141=1 CCC_141=7
BOWEL	Bowel disorder	Dichotomous	0=absence 1=presence .=missing	CCC_171=2 CCC_171=1 CCC_171=7
MIGRAINE	Migraine	Dichotomous	0=absence 1=presence .=missing	CCC_081=2 CCC_081=1 CCC_081=7
CANCER	Cancer	Dichotomous	0=absence 1=presence .=missing	CCC_131=2 CCC_131=1 CCC_131=7
BACKPROB	Back problems excluding fibromyalgia and arthritis	Dichotomous	0=absence 1=presence .=missing	CCC_061=2 CCC_061=1 CCC_061=7

---

## Appendix C – Histograms for the Lognormal and Gamma Distributions of the Study Variables

The graphs for the lognormal and gamma distributions of the variables are shown in Figures 4-18.

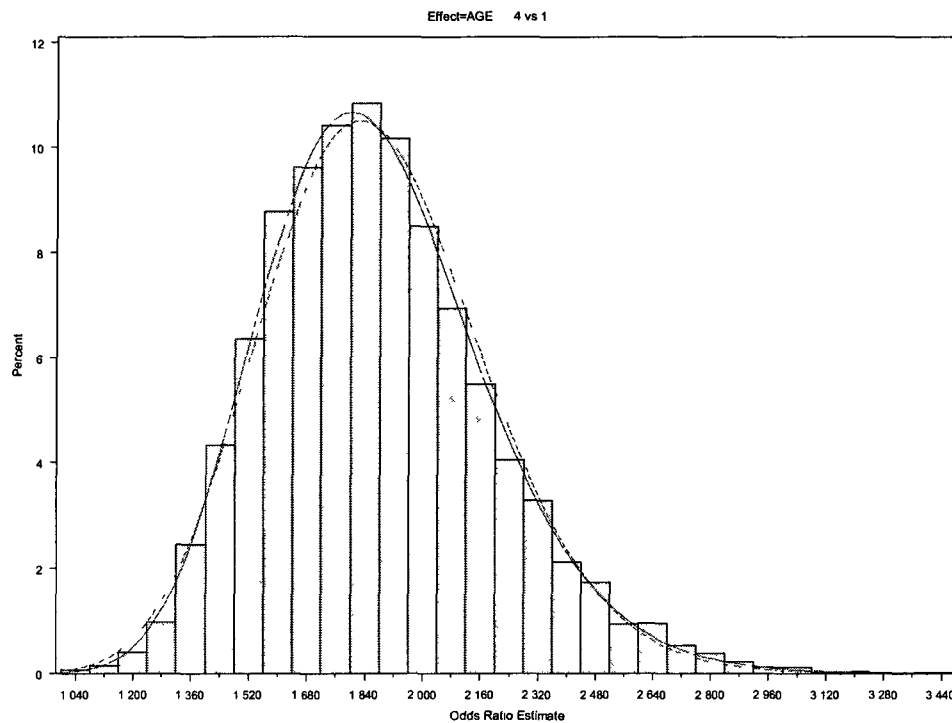


Figure 4. Histogram for AGE 4 vs. 1



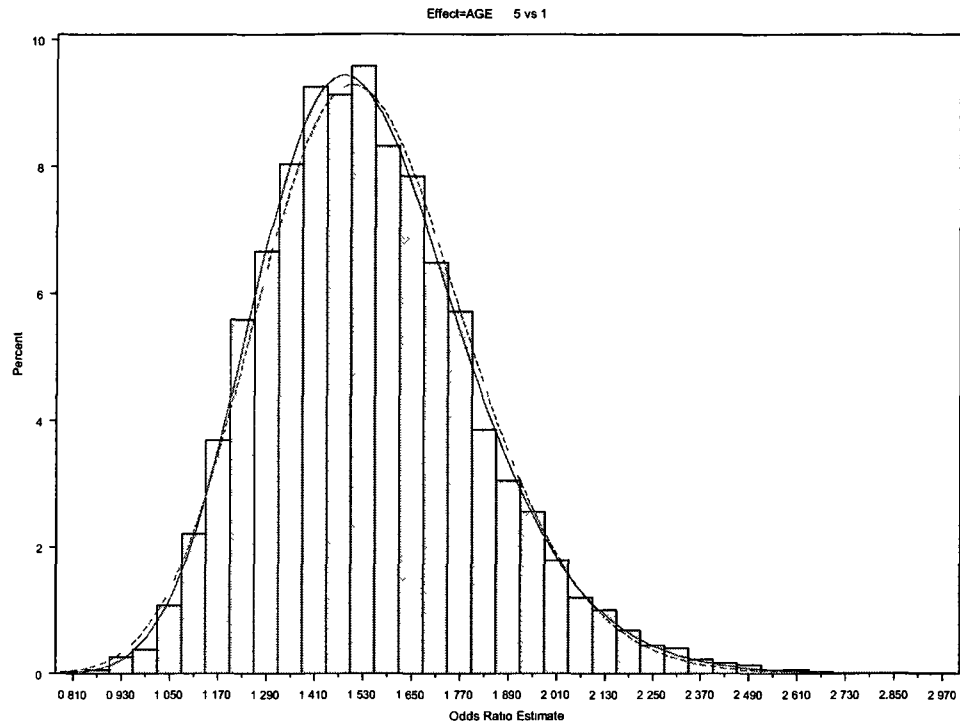


Figure 5. Histogram for AGE 5 vs. 1

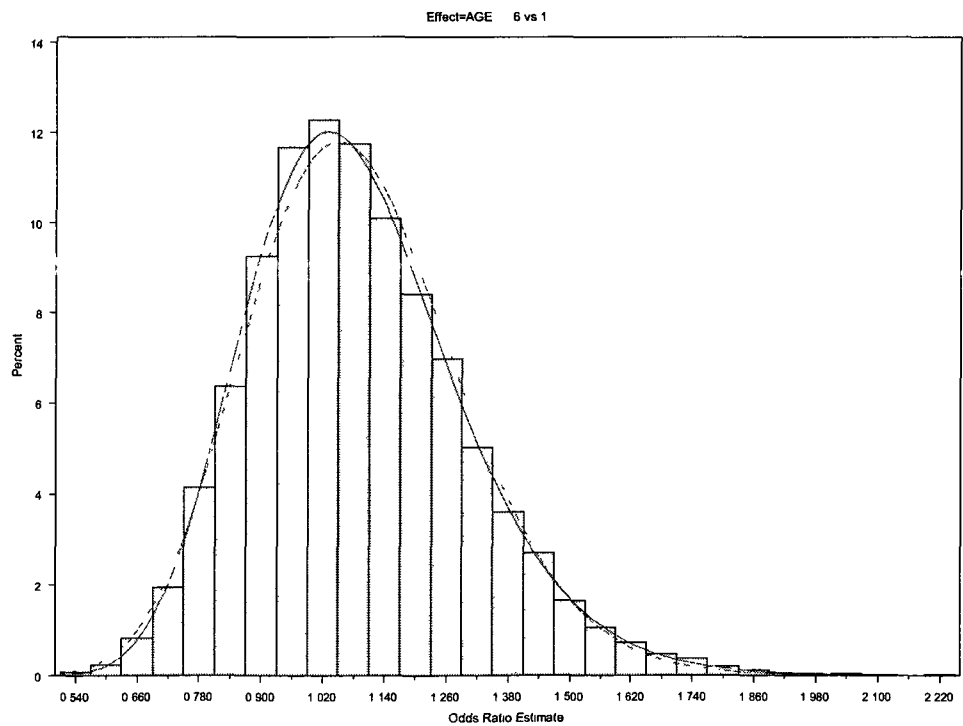


Figure 6. Histogram for AGE 6 vs. 1

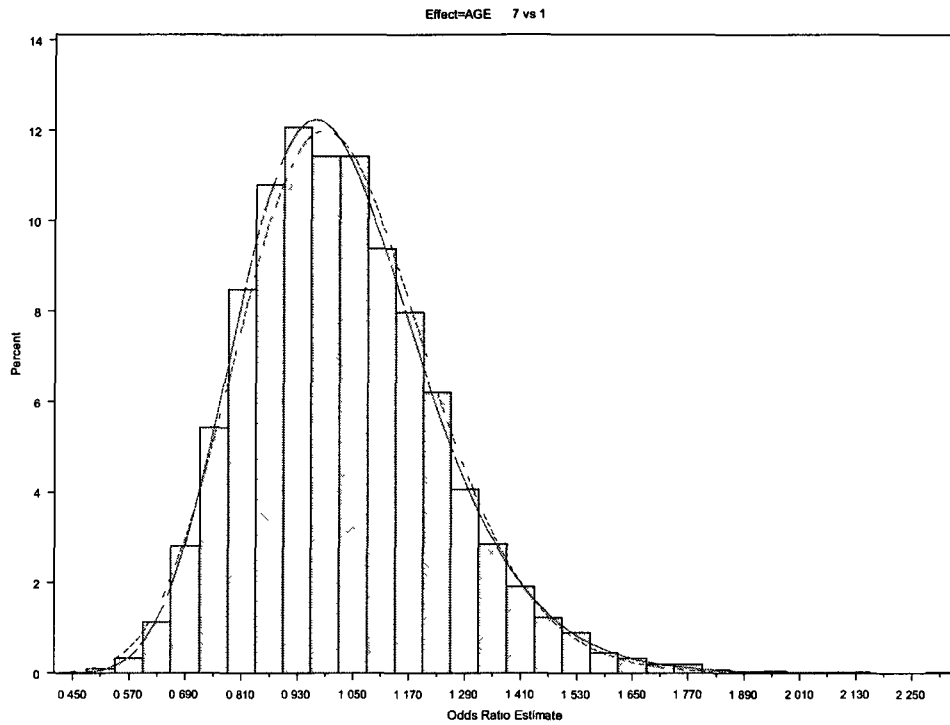


Figure 7. Histogram for AGE 7 vs. 1

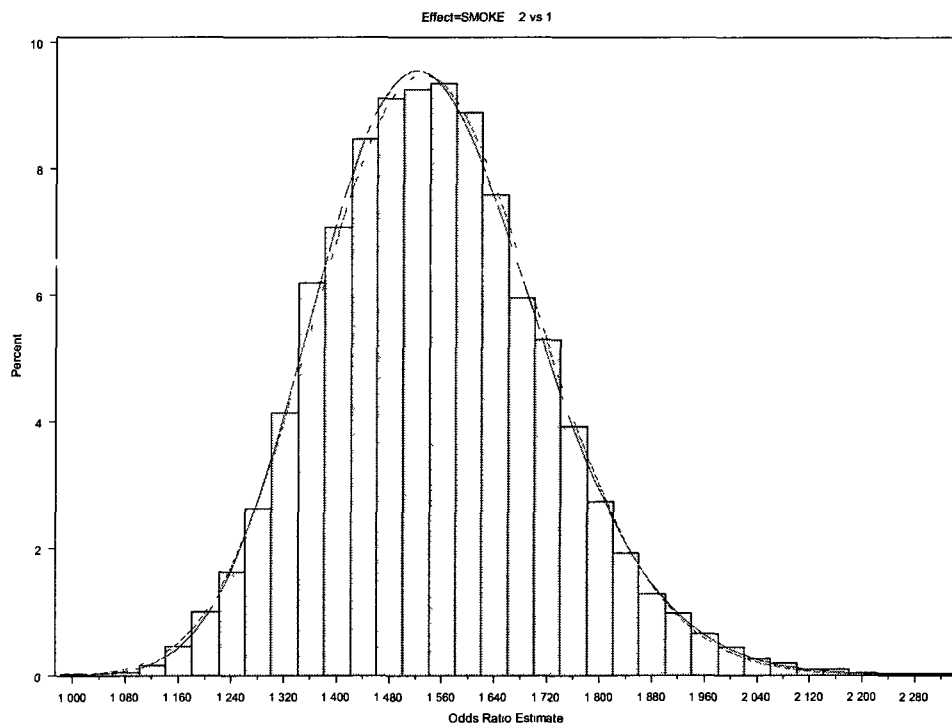


Figure 8. Histogram for SMOKE 2 vs. 1

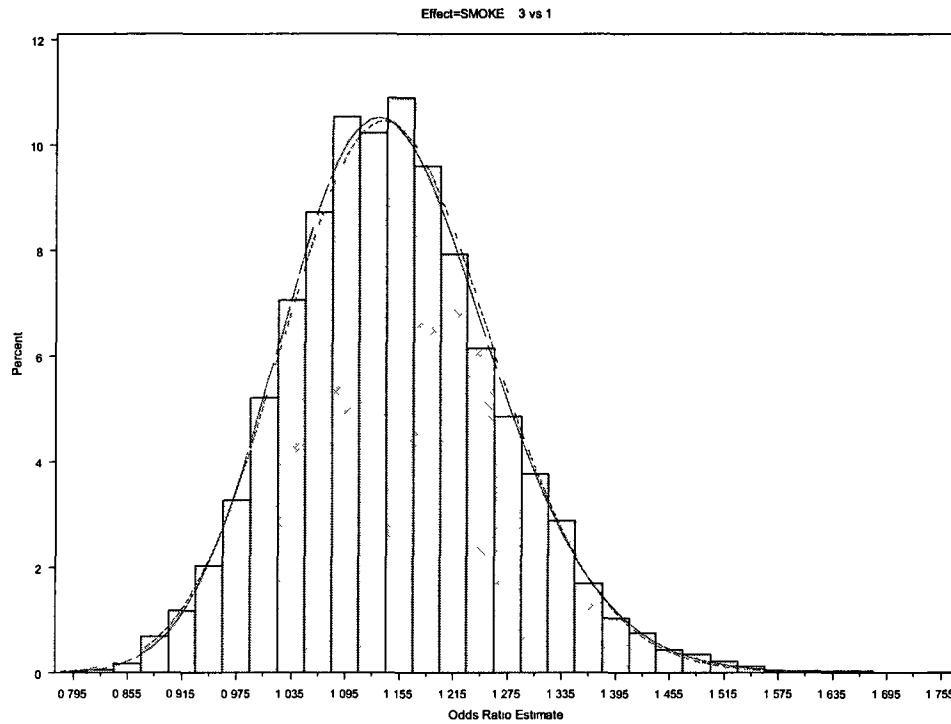


Figure 9. Histogram for SMOKE 3 vs. 1

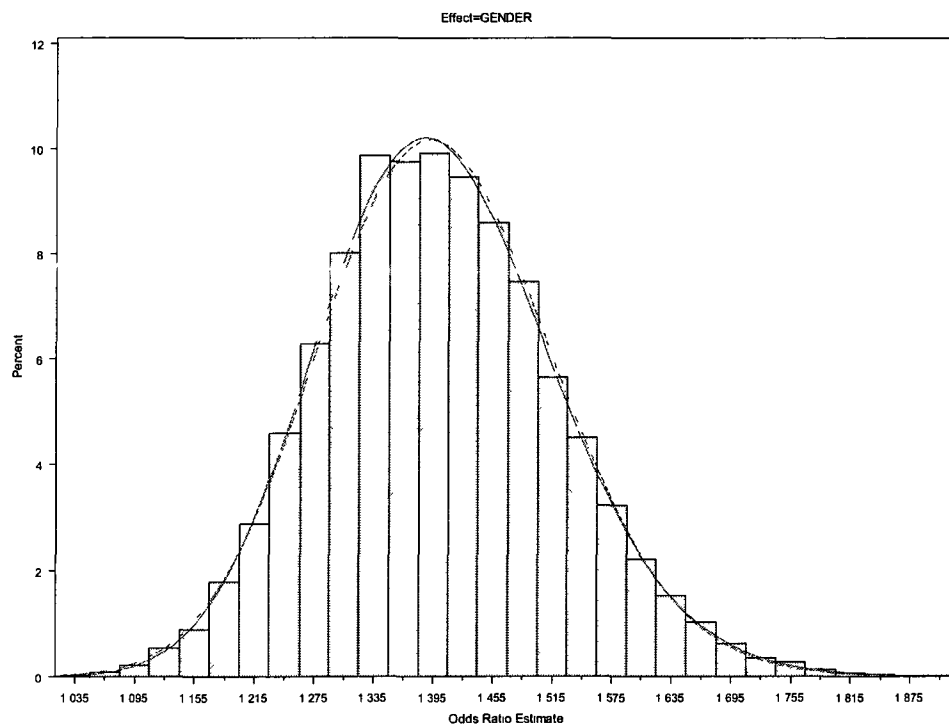


Figure 10. Histogram for GENDER

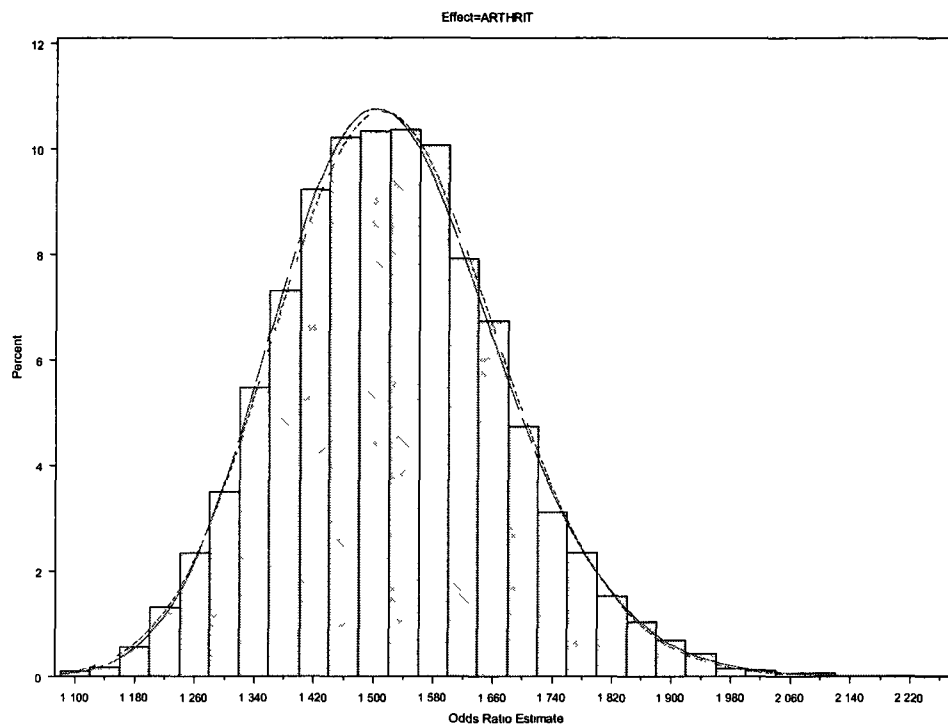


Figure 11. Histogram for ARTHRIT

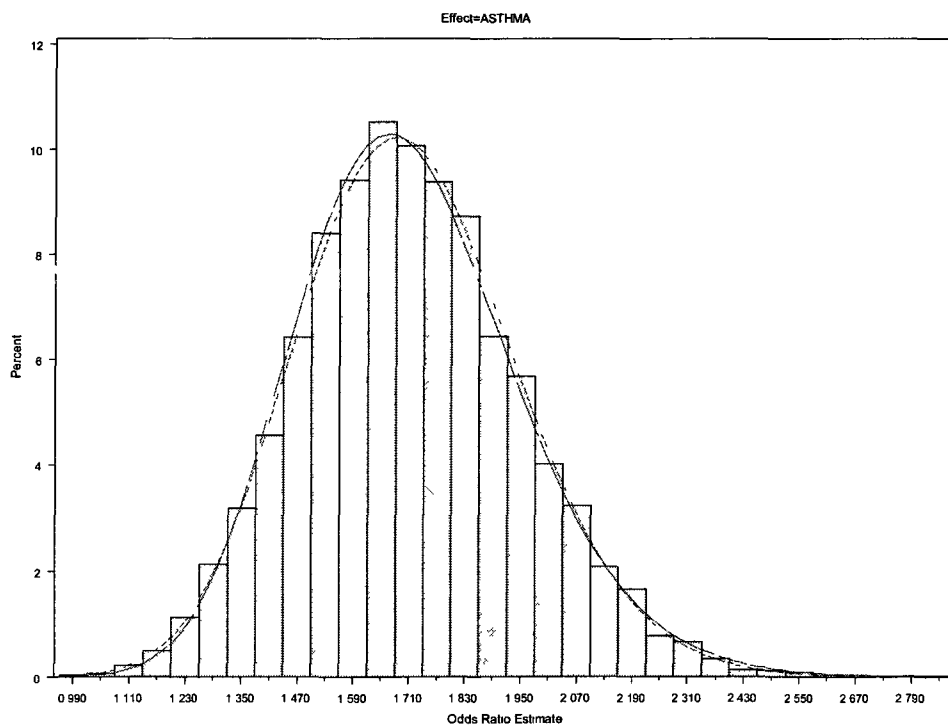


Figure 12. Histogram for ASTHMA

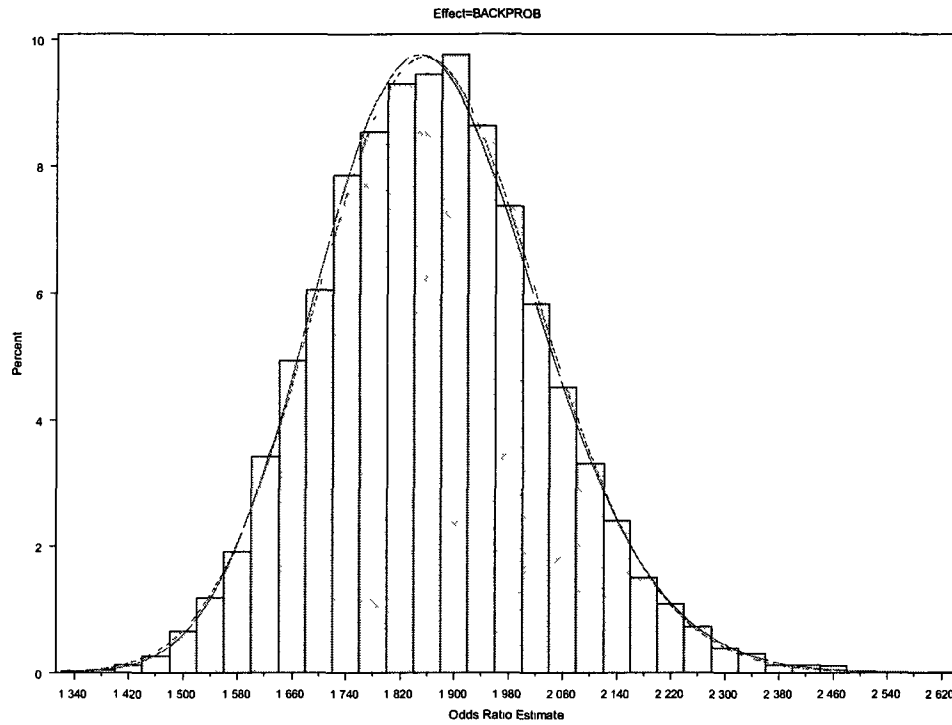


Figure 13. Histogram for BACKPROB

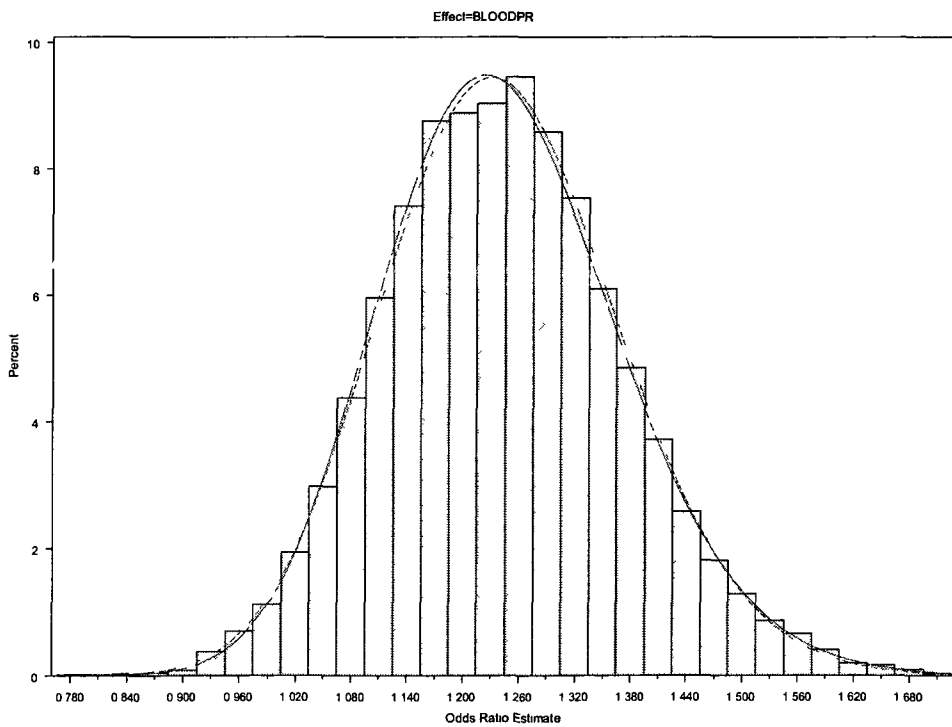


Figure 14. Histogram for BLOODPR

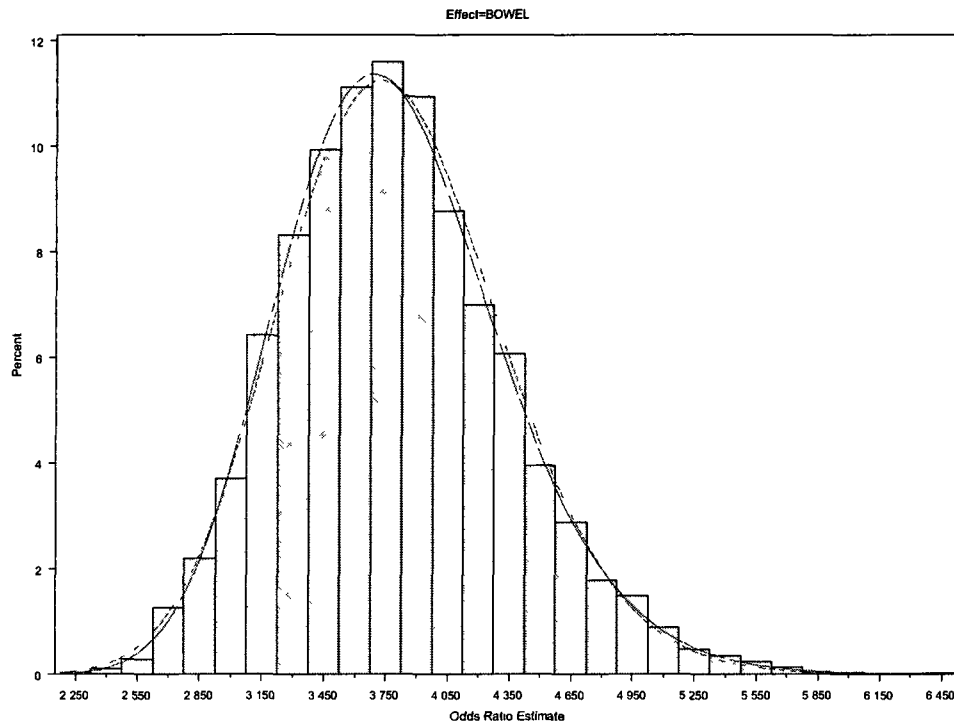


Figure 15. Histogram for BOWEL

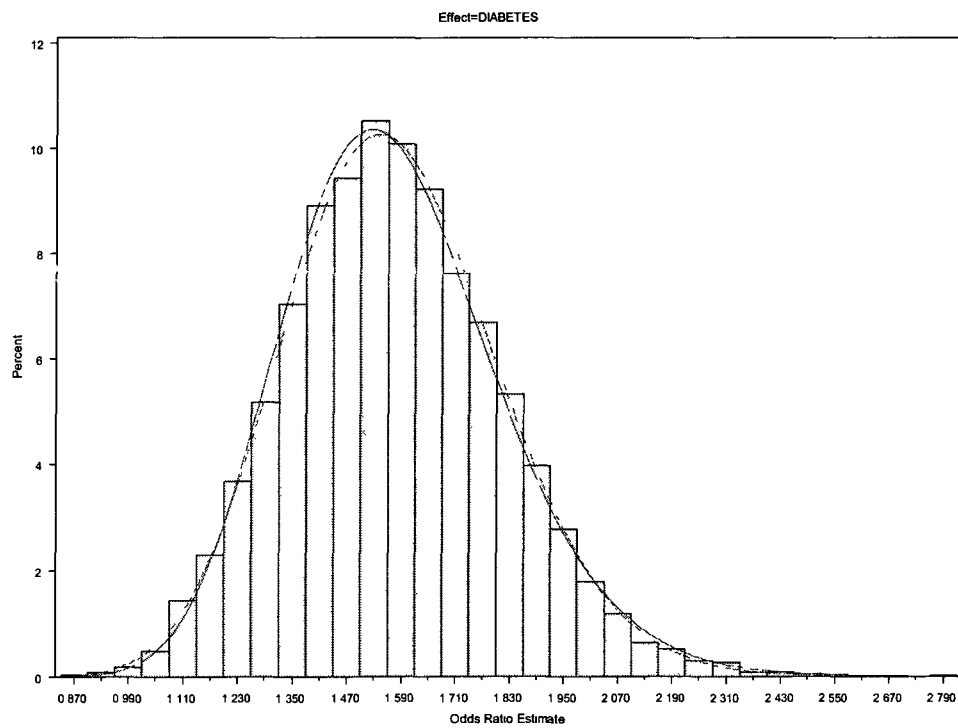


Figure 16. Histogram for DIABETES

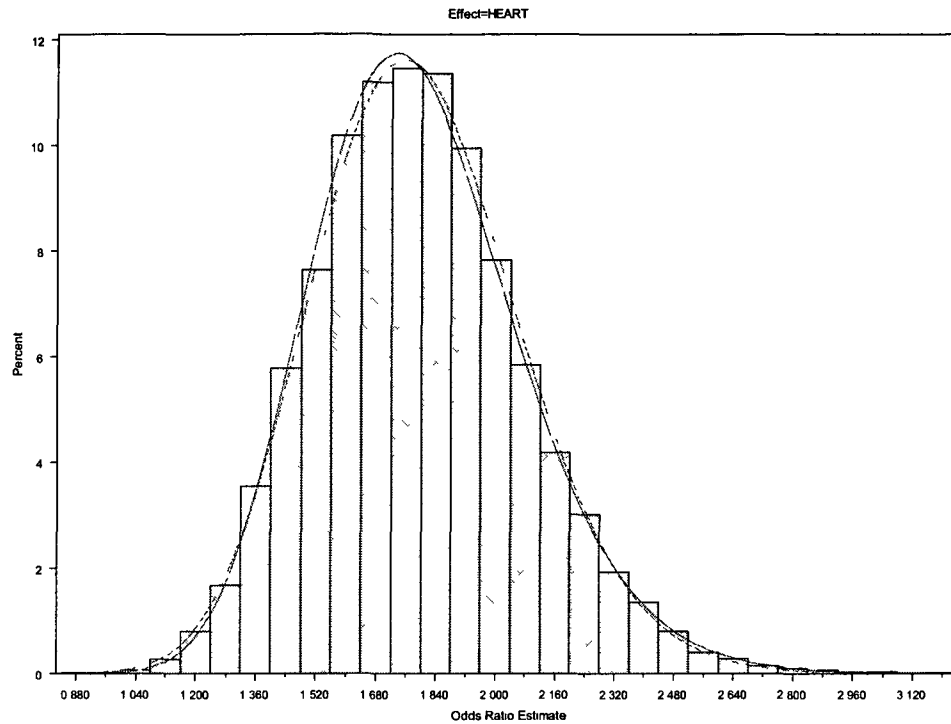


Figure 17. Histogram for HEART

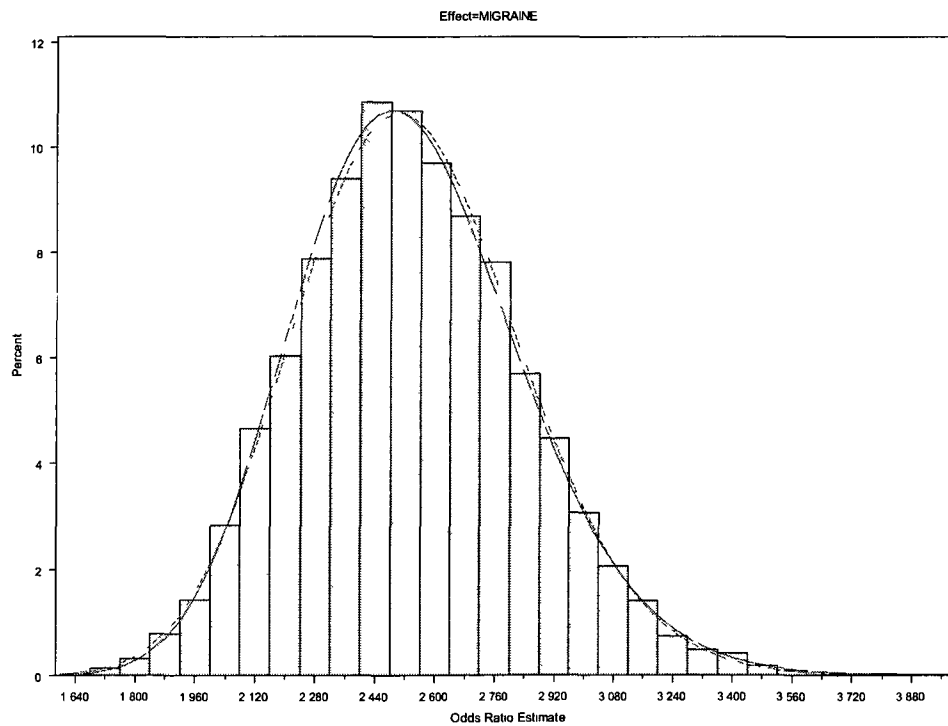


Figure 18. Histogram for MIGRAINE